CIRRELT

Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

# Liquidity-Adjusted Intraday Value at Risk Modeling and Risk Management : An Application to Data from Deutsche Börse

Georges Dionne
Maria Pacurar
Xiaozhou Zhou

February 2014

CIRRELT-2014-11

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

Université Laval · McGill · Concordia · ÉTS · UQÀM · HEC Montréal · Polytechnique Montréal · Université de Montréal

# Liquidity-Adjusted Intraday Value at Risk Modeling and Risk Management: An Application to Data from Deutsche Börse

**Georges Dionne[1,*], Maria Pacurar[2], Xiaozhou Zhou[1]**

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Finance, HEC Montréal, 3000 Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

[2] Department of Finance, Rowe School of Business, Dalhousie University, Kenneth C. Rowe Management Building, 6100 University Avenue, Rowe 2024, Halifax, Canda B3H 4R2

**Abstract.** This paper develops a high-frequency risk measure, the Liquidity-adjusted Intraday Value at Risk (LIVaR). Our objective is to explicitly consider the endogenous liquidity dimension associated with order size. Taking liquidity into consideration when using intraday data is important because significant position changes over very short horizons may have large impacts on stock returns. By reconstructing the open Limit Order Book (LOB) of Deutsche Börse, the changes of tick-by-tick ex-ante frictionless return and actual return are modeled jointly using a Log-ACD-VARMAMGARCH structure. This modeling helps to identify the dynamics of frictionless and actual returns, and to quantify the risk related to the liquidity premium. From a practical perspective, our model can be used not only to identify the impact of ex-ante liquidity risk on total risk, but also to provide an estimation of VaR for the actual return at a point in time. In particular, there will be considerable time saved in constructing the risk measure for the waiting cost because once the models have been identified and estimated, the risk measure over any time horizon can be obtained by simulation without re-sampling the data and re-estimating the model.

**Keywords**. Liquidity-adjusted Intraday Value at Risk (LIVaR), tick-by-tick data, Log-ACD-VARMAMGARCH, ex-ante liquidity premium, limit order book.

**Acknowledgements.** We thank Yann Bilodeau for his help in constructing the dataset and comments. We also thank Diego Amaya, Tolga Cenesizoglu, Philippe Gregoire, Benoit Perron, Gabriel Yergeau and participants at the 53e congrès annuel de la Société canadienne de science économique and the 2013 Canadian Economics Association Conference for their remarks. Georges Dionne and Maria Pacurar acknowledge financial support from the Social Sciences and Humanities Research Council (SSHRC), Xiaozhou Zhou thanks Fond de recherche du Québec - Société et culture (FRQSC) and Centre interuniversitaire sur le risque, les politiques économiques et l'emploi (CIRPEE) for financial support.

Results and views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect those of CIRRELT.

Les résultats et opinions contenus dans cette publication ne reflètent pas nécessairement la position du CIRRELT et n'engagent pas sa responsabilité.

---

* Corresponding author: Georges.Dionne@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2014

© Dionne, Pacurar, Zhou and CIRRELT, 2014

## 1. Introduction

With the help of computerization, many main exchanges around the world such as Euronext, the Tokyo Stock Exchange, Toronto Stock Exchange and the Australian Stock Exchange organize trading activities under a pure automatic order-driven structure: there are no designated market-makers during the continuous trading and the liquidity is fully guaranteed by market participants via an open Limit Order Book (LOB hereafter). In other main exchange markets including NYSE, NASDAQ and Frankfurt Stock Exchange, the trading activities are carried out under the automatic order-driven structure and the traditional floor-based quote-driven structure. Nevertheless, most trades are executed under the automatic order-driven structure due to its advantage of transparency, efficiency and immediacy. Consequently, the frequency of trading is becoming shorter and trading activity has become easier than ever before.

One type of trading behavior is active trading or day-trading by which traders trade during the day and liquidate all open positions before market closing. Besides high-frequency traders, financial institutions also need intraday risk analysis (Gouriéroux and Jasiak (2010)) for internal control of their trading desks. As a result, an active trading or day-trading culture requires institutional investors, active individual traders and even regulators of financial markets to pay more and more attention to intraday risk management. [1] However, traditional risk management has been challenged by this trend towards high-frequency trading because the low-frequency measures of risk such as Value at Risk (VaR), which is usually based on daily data, struggle to capture the potential liquidity risk hidden in very short horizons. Typically, the well recognized description of a liquid security is that there is the ability to convert the *desired quantity* of the financial asset into cash *quickly* and with little *impact on the market price* (Demsetz (1968); Black (1971); Kyle (1985); Glosten and Harris (1988)). Four dimensions are implicitly included in this definition: *volume*

[1] High-frequency trading was estimated to make up 51% of equity trades in the U.S. in 2012 and 39% of traded value in the European cash markets (Tabb Group).

(significant quantity), *price impact* (deviation from the best price provided in the market), *time* (speed to complete the transaction) and *resilience* (speed to backfilling). In an automated order-driven trading system, because the liquidity is fully provided by the open LOB, we can investigate liquidity risk by monitoring the evolution of LOB and exploring the corresponding embedded information.

The traditional VaR can be interpreted as estimation of the potential loss on a predetermined portfolio over a relatively long fixed period, that is, a measure of price risk. This is not a liquidation value since it does not take into account the volume dimension but solely a 'paper value' for a frozen portfolio. However, liquidity risk is always present when the transaction is not yet realized. Furthermore, from a high-frequency market microstructure perspective, the transaction price is an outcome of information shock, trading environment, market imperfections and the state of the LOB. For very short horizons, all these microstructure effects could cause the transaction price to deviate from the efficient price. Therefore, if we concentrate on the actual liquidation value resulting from an active trading operation, the traditional VaR may be characterized by a serious omission of liquidity, especially when the liquidation quantity is large. To make the VaR measure more accurate in evaluating the liquidation value, one should include an additional dimension of risk, the ex-ante liquidity risk. From a practical perspective, this introduction of a liquidity dimension can offer a more accurate risk measure for intraday or active traders and market regulators who aim to closely monitor the total risk of the market.

Few studies have focused on high-frequency risk measures. Dionne, Duchesne and Pacurar (2009) are the first to consider a ultra-high-frequency market risk measure, Intraday Value at Risk (IVaR) based on *all* transactions. In their study, the informative content of trading frequency is taken into account by modeling the durations between two consecutive transactions. One important practical contribution is that, instead of being restricted to traditional one- or five-minute horizons, their model allows for computing the IVaR measure for any horizon.

The authors found that ignoring the effect of durations can underestimate risk. However, as they noted, similar to other VaR measures, the IVaR ignores the ex-ante liquidity dimension by only taking into account information about transaction prices.

On the other hand, in spite of the rapid development of electronic trading platforms, high-frequency risk measures hardly incorporate the endogenous relationship between open LOB dynamics and frictionless prices, although some recent theoretical and empirical works find that the LOB is informative about the volatility of efficient price (Pascual and Veredas (2010)), and that the volume-volatility relationship is related negatively to the LOB slope (Næs and Skjeltorp (2006)). One exception is Giot and Grammig (2006) who consider the ex-ante liquidity provided by the LOB using data from Xetra, an automated auction system. They focus on the ex-ante liquidity risk faced by an impatient trader acting as a liquidity demander by submitting a market order. The quantity to liquidate is fictive and the choice of quantity is motivated by descriptive statistics and trading statistics of the underlying stock. The use of open LOB data allows them to construct an actual return that is a potential implicit return for a predetermined volume to trade over a fixed time horizon of 10 or 30 minutes. The ex-ante liquidity risk is then quantified by comparing the standard VaR based on frictionless return, i.e. mid-quote return, and the liquidity-adjusted VaR inferred from the actual return. Since the proposed VaR is computed on data from regularly spaced intervals of 10 or 30 minutes and the snapshot of the LOB is taken at the beginning and end of the interval, all the information within the interval is omitted.

The development of liquidity-adjusted risk measures goes back to Bangia et al. (1999) who first consider the liquidity risk in VaR computation. Their liquidity risk is measured by the half spread, which is furthermore assumed to be uncorrelated with market risk. Actually, the total risk they attempt to identify is the sum of market risk and trading cost associated to only one share since their liquidity risk

measure does not consider the volume dimension. In line with Bangia et al. (1999), Angelidis and Benos (2006) estimate the liquidity-adjusted VaR by using data from the Athens Stock Exchange. They find that liquidity risk measured by the bid-ask spread accounts for 3.4% for high-capitalization stocks and 11% for low-capitalization stocks. Using the same framework as Bangia et al. (1999), Weiβ and Supper (2013) address liquidity risk of a five-NASDAQ stock portfolio by estimating the multivariate distribution of both log-return and spread using the vine copulas to account for the dependence between the two across firms over a regular interval of 5 minutes. They evidence strong extreme comovements in liquidity and tail dependence between bid-ask spreads and log-returns across the selected stocks.

Our paper is related to the market microstructure literature that investigates the role of high-frequency liquidity risk. However, our study differs from previous papers in several ways. First, we examine the ex-ante liquidity risk by focusing on the tick-by-tick frictionless return and actual returns derived from open LOB. By ex-ante we mean that the transaction does not really occur but we can still obtain a potential transaction price issued from a predetermined volume to trade if we have information from the reconstructed LOB. Most papers in the literature are interested in ex-post liquidity, which is already consumed by the market or marketable orders when the trades complete. Compared to ex-post liquidity, ex-ante liquidity is more informative and relevant as it measures the unconsumed liquidity in LOB. Besides, the existing literature that analyzes ex-ante measures of liquidity (e.g., Giot and Grammig, 2006) is based on a regular time interval and thus ignores the information during the intervals. In our study, we evidence that the durations between two consecutive observations do have a positive relation with volatilities of actual returns and frictionless returns.

Second, our study addresses the questions of what is the relationship between liquidity risk and market risk and how does the ex-ante liquidity evolve during the trading day. There are several empirical papers that attempt to identify the

proportion of risk associated with liquidity as discussed above. However, to the best of our knowledge, their identified liquidity is ex-post liquidity without studying the dynamics of ex-ante liquidity derived from LOB and its relation with market risk. Nevertheless, one challenge of directly modeling frictionless returns and actual returns (issued from the best bid/ask price and potential liquidation price, respectively) is that they are not time-additive. Therefore, in this paper, we model the frictionless return changes and actual return changes using an econometric system characterized by the Logarithmic Autoregressive Conditional Duration, Vector Autoregressive Moving Average and Multivariate GARCH processes (denoted by Log-ACD, VARMA and M-GARCH hereafter). The structure will not only capture the joint dynamics of both frictionless return changes and actual return changes, but also quantify the impact of ex-ante liquidity risk on total risk by further defining $IVaR^c$ and $LIVaR^c$ as the VaRs on frictionless return changes and actual return changes, respectively. In order to make the model more flexible, we allow for the time-varying correlation of volatility of the frictionless return and actual return.

Third, from a practical perspective, our proposed risk measure aims at providing a global view of ex-ante liquidity that can help high-frequency traders develop their timing strategies during a particular trading day. Our model is first estimated on deseasonalized data and then validated on both simulated deseasonalized and re-seasonalized data. The time series of re-seasonalized data is constructed by re-introducing the deterministic seasonality factors. One advantage of simulated re-seasonalized data is that risk management can be conducted under calendar time. In addition, as the model is estimated using tick-by-tick observations and takes into account the durations between two consecutive transactions, practitioners can construct the risk measure for any desired time horizon.

The rest of the paper is organized as follows: Section 2 describes the Xetra trading system and the dataset we utilize. Section 3 briefly presents the procedure used to test the model and compute Impact coefficient of ex-ante liquidity risk and Ex-

ante liquidity premium. Section 4 defines the actual return, frictionless return, IVaR and LIVaR. In Section 5, we present the econometric model used to capture the dynamics of duration, frictionless return changes, actual return changes and their correlation. Section 6 applies the econometric model proposed in Section 5 to data for stocks RWE AG, Merck and SAP and reports the estimation results. The model performance is assessed by the Unconditional Coverage test of Kupiec (1995) and the Independence test proposed by Christoffersen (1998). In Section 7, we discuss the ex-ante liquidity risk in $LIVaR^c$ and compare the proposed LIVaR with other high-frequency risk measures. Section 8 concludes and provides possible new research directions.

## 2. Xetra trading system and dataset

As mentioned above, electronic trading systems have gained popularity in many stock exchanges during the two last decades. The present study uses data from the automated order-driven trading system Xetra, which is operated by Deutsche Börse at Frankfurt Stock Exchange (FSE) and has a similar structure to Integrated Single Book of NASDAQ and Super Dot of NYSE. It is the main trading platform in Germany and realizes more than 90% of total transactions at German exchanges. The trading and order processing (entry, revision, execution and cancellation) of the Xetra system are highly computerized and maintained by the German Stock Exchange. Since September 20, 1999, trading hours have been from 9h00 to 17h30 CET (Central European Time). However, during the pre- and post-trading hours, operations such as entry, revision and cancellation are still permitted.

To ensure trading efficiency, Xetra operates with different market models that define order matching, price determination, transparency, etc. One of the important parameters of a market model is the trading model that determines whether the trading is organized in a continuous or discrete way or both. During normal trading hours, there are two types of trading mechanisms: call auction and

continuous auction. Call auction could occur one or several times during the trading day in which the clearance price is determined by the state of LOB and remains as the open price for the following continuous auction. Furthermore, at each call auction, market participants can submit both round-lot and odd-lot orders, and both start and end time for a call auction are randomly chosen by a computer to avoid scheduled trading. For the stocks in DAX30 index, there are three auctions during a trading day—the open, mid-day and closing auctions. The mid-day auction starts at 13h00 and lasts around 2 minutes. Between the call auctions, the market is organized as a continuous auction where traders can only submit round-lot-sized limit orders or market orders.

For blue chip and other highly liquid stocks, during the continuous trading there are no dedicated market makers like the traditional NYSE specialists. Therefore, the liquidity comes from all market participants who submit limit orders in LOB. In the Xetra trading system, most of the market models impose the Price-Time-Priority condition where the electronic trading system places the incoming order after checking the price and timestamps of all available limit orders in LOB. Our database includes up to 20 levels of LOB information except the hidden part of an iceberg order, which means that by observing the LOB, any trader and registered member can monitor the dynamic of liquidity supply and potential price impact caused by a market or marketable limit order. However, all the trading and order submission are anonymous, that is, the state and the updates of LOB can be observed but there is no information on the identities of market participants.

The raw dataset that we have access to contains all the events that are tracked and sent through the so-called data streams. There are two main types of streams: delta and snapshot. The former tracks all the possible updates in LOB such as entry, revision, cancellation and expiration. Traders can be connected to the delta stream during the trading hours to receive the latest information, whereas the second aims at giving an overview of the state of LOB and is sent after a constant time interval

for a given stock. Xetra original data with delta and snapshot messages are first processed using the software XetraParser developed by Bilodeau (2013) in order to make Deutsche Börse Xetra raw data usable for academic and professional purposes. XetraParser reconstructs the real-time order book sequence including all the information for both auctions and continuous trading by implementing the Xetra trading protocol and Enhanced Broadcast.[2] We further convert the raw LOB information into a readable LOB for each update time and then retrieve useful and accurate information about the state of LOB and the precise timestamp for order modifications and transactions during the continuous trading. Inter-trade durations as well as LOB update durations are irregular. The stocks SAP (SAP), RWE AG (RWE) and Merck (MRK) that we choose for this study are blue chip stocks from the DAX30 index. SAP is a leading multinational software corporation with a market capitalization of 33.84 billion Euros in 2010. RWE generates and distributes electricity to various customers including municipal, industrial, commercial and residential customers. The company produces natural gas and oil, mines coal and delivers and distributes gas. In 2010, its market capitalization was around 15 billion Euros. Merck is the world's oldest operating chemical and pharmaceutical company with a market capitalization of 4 billion Euros in 2010.

## 3. Procedure used for computing the risk measures

To compute the proposed risk measures, the model will be first estimated using deseasonalized data and then the tests will be carried out on both deseasonalized and seasonalized data. We present the flowchart where we illustrate the steps to follow. More precisely, our study proceeds as follows:

[Insert Flowchart here]

[2] See SolutionXetra Release 11.0 – Enhanced Broadcast solution and Interface specification for a detailed description.

(a) We first compute the raw tick-by-tick durations defined as the time interval between two consecutive trades and tick-by-tick frictionless returns and actual returns based on the data of open LOB and trades.

(b) We further compute the frictionless return changes and actual return changes by taking the first difference of frictionless returns and actual returns, respectively. This step is required because the frictionless return and actual return are not time-additive and cannot be modeled directly.

(c) We remove seasonality from durations, frictionless return changes and actual return changes to obtain the corresponding deseasonalized data.

(d) The deseasonalized data are modeled by the LogACD-VARMA-MGARCH model.

(e) Once we estimate the model, we simulate the deseasonalized data based on estimated coefficients and construct VaR measures at different confidence levels for backtesting on out-of-sample deseasonalized data.

(f) The seasonal factors are re-introduced into deseasonalized data to generate the re-seasonalized data.

(g) As done in (e), we construct different quantiles for backtesting on out-of-sample seasonalized data.

(h) We construct Impact coefficients of ex-ante liquidity risk based on the simulated re-seasonalized data.

(i) We further compute the IVaR and LIVaR, which are defined as the VaR for frictionless return and actual return, respectively.

(j) Based on IVaR and LIVaR, we can finally compute the ex-ante liquidity premium by taking the ratio of the difference between IVaR and LIVaR over LIVaR.

The following sections will explain each step in detail.

## 4. Frictionless return, actual return and the corresponding high-frequency VaRs

We take into account all information available in the continuous auction by modeling tick-by-tick data. The first characteristic in tick-by-tick data modeling is that the durations between two consecutive transactions are irregularly spaced.

Consider two consecutive trades that arrive at $t_{i-1}$ and $t_i$, and define $dur_i$ as the duration from $t_{i-1}$ to $t_i$. Based on this point process, we can further construct two return processes. One is frictionless return and the other is actual return. More specifically, the frictionless return is defined as the log ratio of best bid price, $b_i(1)$ at moment $i$ and previous best ask price, $a_{i-1}(1)$. The frictionless return is an ex-ante return indicating the tick-by-tick return for selling only one unit of stock.

$$R_i^F = \ln(\frac{b_i(1)}{a_{i-1}(1)}) \tag{1}$$

The actual return is defined as the log ratio of selling price for a volume $v$ and previous best ask price.[3]

$$R_i^B = \ln(\frac{b_i(v)}{a_{i-1}(1)})$$

$$\text{where } b_i(v) = \frac{\sum_{k=1}^{K-1} b_{k,i} v_{k,i} + b_{K,i} v_{K,i}}{v} \quad \text{and} \quad v_{K,i} = v - \sum_{k=1}^{K-1} v_{k,i} \tag{2}$$

$b_{k,i}$ *and* $v_{k,i}$ are the $k$th level bid price and volume available, respectively. $v_{K,i}$ is the quantity left after K-1 levels are completely consumed by $v$. The consideration of quantity available in LOB is in line with other ex-ante liquidity measures in market microstructure literature (Irvine, Benston and Kandel (2000); Domowitz, Hansch and Wang (2005); Coppejans, Domowitz and Madhavan (2004), among others). The choice of volume $v$ is motivated by transaction volume and volume available in LOB. Explicitly, for each stock we first compute the cumulative volume available over the 20 levels at each transaction moment for the bid side of LOB, and then we choose the minimum cumulative volume as the maximum volume to construct the actual returns. By doing so, we avoid the situation where the actual price does not exist for a given volume. One concern that may arise involves the iceberg orders, which keep a portion of quantity invisible to the market

[3] The frictionless return and actual return from the point of view of buyers can be defined similarly.

participants. In this study, we assume that the liquidity risk is faced by an impatient trader and the possibility of trading against an iceberg order will not influence his trading behavior. Furthermore, as noted by Beltran-Lopez, Giot and Grammig (2009), the hidden part of the book does not carry economically significant informational content. The difference between frictionless return and actual return is that actual return takes into account the desired transaction volume, which is essential for the liquidity measure. Intuitively, the actual return measures the ex-ante return when liquidating $v$ units of shares.

One characteristic of our defined frictionless returns and actual returns is that they do not possess the time-additivity property as traditional log-returns. To circumvent this difficulty, we model the frictionless return changes and actual return changes instead of modeling the actual return and frictionless return directly. More specifically, let $r_t^f = R_t^F - R_{t-1}^F$ and $r_t^b = R_t^B - R_{t-1}^B$ be the tick-by-tick frictionless return changes and actual return changes. Following this setup, the L-step forward frictionless return and actual return can be expressed as follows:

$$R_{t+L}^F = \ln(\frac{b_{t+L}(1)}{a_{t+L-1}(1)}) = \ln(\frac{b_t(1)}{a_{t-1}(1)}) + r_{t+1}^f + r_{t+2}^f + \ldots r_{t+L}^f = R_t^F + \sum_{m=1}^{L} r_{t+m}^f \quad (3)$$

$$R_{t+L}^B = \ln(\frac{b_{t+L}(v)}{a_{t+L-1}(1)}) = \ln(\frac{b_t(v)}{a_{t-1}(1)}) + r_{t+1}^b + r_{t+2}^b + \ldots r_{t+L}^b = R_t^B + \sum_{m=1}^{L} r_{t+m}^b \quad (4)$$

The terms $\sum_{m=1}^{L} r_{t+m}^f$ and $\sum_{m=1}^{L} r_{t+m}^b$ are the sum of all tick-by-tick changes in return over a predetermined interval and can be considered as the waiting cost related to frictionless returns and actual return. [4] More specifically, they measure the costs/gains associated with the latter instead of immediate liquidation of 1 share and $v$ shares of stock, respectively. Moreover, if we define the $1VaR^c$ and $LIVaR^c$ as the VaR for $\sum_{m=1}^{L} r_{t+m}^f$ and $\sum_{m=1}^{L} r_{t+m}^b$, respectively and compute them from a

[4] The waiting cost could be positive or negative, which indicates loss and gain, respectively.

predetermined time interval, then the IVaR$^c$ and LIVaR$^c$ will provide the possible loss over a given interval for an investor that trades at frictionless or actual return. In other words, the IVaR$^c$ and LIVaR$^c$ will estimate the possible loss in terms of frictionless return and actual return, which are related to market risk and total risk (market risk and ex-ante liquidity risk), respectively. Mathematically, consider a realization of a sequence of intervals with length *int* and let $y^f_{int,t}$ *and* $y^b_{int,t}$ [5] be the sum of tick-by-tick changes of returns $r^f_i$ *and* $r^b_i$ over *t*-th interval,

$$y^f_{int,t} = \sum_{j=\tau(t-1)}^{\tau(t)-1} r^f_j \quad ; \quad y^b_{int,t} = \sum_{j=\tau(t-1)}^{\tau(t)-1} r^b_j \tag{5}$$

where $\tau(t)$ is the index for which the cumulative duration exceeds *t*-th interval with length *int* for the first time. By definition

$$int \geq \sum_{j=\tau(t-1)}^{\tau(t)-1} dur_j \text{ and } int \leq \sum_{j=\tau(t-1)}^{\tau(t)} dur_j. \tag{6}$$

The process of duration allows aggregating the tick-by-tick data to construct the dynamic of frictionless return and actual return for a predetermined interval that will allow for consideration of risk in calendar time. Accordingly, the IVaR$^c$ and LIVaR$^c$ [6] for frictionless return changes and actual return changes with confidence level $1-\alpha$ for a predetermined interval *int* are defined as

$$Pr(y^f_{int,t} < IVaR^c_{int,t}(\alpha) | I_t) = \alpha$$

$$Pr(y^b_{int,t} < LIVaR^c_{int,v,t}(\alpha) | I_t) = \alpha$$

[5] $y^f_{int}$ *and* $y^b_{int}$ are defined on both seasonalized return changes and deseasonalized return changes.

[6] IVaR$^c$ and LIVaR$^c$ are also defined on both seasonalized return changes and deseasonalized return changes. Moreover, our IVaR$^c$ and LIVaR$^c$ can also be used in the strategy of short selling where IVaR$^c$ and LIVaR$^c$ will be the 1-α quantiles of the distributions.

$I_t$ is the information set until moment $\tau(t-1)$. Similar to the traditional definition of VaR, $IVaR^c_{int,t}(\alpha)$ and $LIVaR^c_{int,v,t}(\alpha)$ are the conditional α-quantiles for $y^f_{int,t}$ and $y^b_{int,t}$.

We can further define the IVaR and LIVaR as the VaR for the frictionless return and actual return as following:

$$IVaR_{int,t} = R^F_{\tau(t-1)} + IVaR^c_{int,t}$$

$$LIVaR_{int,v,t} = R^B_{\tau(t-1)} + LIVaR^c_{int,v,t}$$

where $R^F_{\tau(t-1)}$ and $R^B_{\tau(t-1)}$ are the frictionless return and actual return at the beginning of the $t$-th interval. Consequently, IVaR and LIVaR estimate the α-quantiles for frictionless return and actual return at the end of the $t$-th interval.

## 5. Methodology

In our tick-by-tick modeling, there are three random processes: the duration, the changes of frictionless return, and the changes of actual return. The present study assumes that the duration evolution is strongly exogenous but has an impact on the volatility of frictionless and actual return changes. The joint distribution of duration, frictionless return change and actual return change can be decomposed into the marginal distribution of duration and joint distribution of frictionless and actual return change conditional on duration. More specifically, the joint distribution of the three variables is:

$$\begin{aligned} &f^{d,f,b}(dur_i, r_i^f, r_i^b \,|\, dur_{i-1}, r_{i-1}^f, r_{i-1}^b; \Theta^f) \\ &= f^d(dur_i \,|\, dur_{i-1}, r_{i-1}^f, r_{i-1}^b; \theta^d) f^{f,b}(r_i^f, r_i^b \,|\, dur_i, dur_{i-1}, r_{i-1}^f, r_{i-1}^b; \theta^d, \theta^{f,b}) \end{aligned} \quad (7)$$

where $f^{d,f,b}$ is the joint distribution for duration, frictionless return change and actual return change. $f^d(\cdot)$ is the marginal density for duration and $f^{f,b}(\cdot)$ is the

joint density for actual and frictionless return changes. Consequently, the corresponding log-likelihood function for each joint distribution can be written as:

$$L(\theta^d, \theta^{f,b}) = \sum_{i=1}^{n} \log f^d(dur_i \left| dur_{i-1}, r_{i-1}^f, r_{i-1}^b; \theta^d \right.) + \log f^{f,b}(r_i^f, r_i^b \left| dur_{i-1}, r_{i-1}^f, r_{i-1}^b, dur_i; \theta^{f,b} \right.) \quad (8)$$

In the next subsections, we specify marginal density for dynamics of duration and joint density for frictionless return and actual return changes. We present the model for deseasonalized duration, frictionless and actual return changes. The deseasonalization procedure is described in detail in Section 6.

5.1 Model for duration

The ACD model used to model the duration between two consecutive transactions was introduced by Engle and Russell (1998). The GARCH-style structure is introduced to capture the duration clustering observed in high-frequency financial data. The basic assumption is that the realized duration is driven by its conditional duration and a positive random variable as error term. Let $\psi_i = E(dur_i \left| I_{i-1} \right.)$ be the expected duration given all the information up to *i-1*, and $\varepsilon_i$ be the positive random variable. The duration can be expressed as: $dur_i = \psi_i \cdot \varepsilon_i$. There are several possible specifications for the expected duration and the independent and identically distributed (i.i.d.) positive random error (see Hautsch (2004) and Pacurar (2008) for surveys). In order to guarantee the positivity of duration, we adopt the log-ACD model proposed by Bauwens and Giot (2000). The specification for expected duration is

$$\psi_i = \exp\left( \omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{i-j} + \sum_{j=1}^{q} \beta_j \ln \psi_{i-j} \right). \quad (9)$$

For positive random errors, we use the generalized gamma distribution, which allows a non-monotonic hazard function and nests the Weibull distribution (Grammig and Maurer (2000); Zhang, Russell and Tsay (2001)):

$$p(\varepsilon_i|\gamma_1,\gamma_2)=\begin{cases}\dfrac{\gamma_1\varepsilon_i^{\gamma_1\gamma_2-1}}{\gamma_3^{\gamma_1\gamma_2}\Gamma(\gamma_2)}\exp[-\left(\dfrac{\varepsilon_i}{\gamma_3}\right)^{\gamma_1}], \varepsilon_i>0, \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{,otherwise.}\end{cases} \tag{10}$$

where $\gamma_1,\gamma_2>0$, $\Gamma(.)$ is the gamma function, and $\gamma_3=\Gamma(\gamma_2)/\Gamma(\gamma_2+\frac{1}{\gamma_1})$.

5.2 Model for frictionless return and actual return changes

The high-frequency frictionless return and actual return changes display a high serial correlation. To capture this microstructure effect, we follow Ghysels and Jasiak (1998) and adopt a VARMA (*p*,*q*) structure:

$$R_i=\left(r_i^b,r_i^f\right)',\ E_i=\left(e_i^b,e_i^f\right)'$$

$$R_i=\sum_{m=1}^{p}\Phi_m R_{i-m}+E_i-\sum_{n=1}^{q}\Delta_n E_{i-n} \tag{11}$$

where $\Phi_m$ and $\Delta_n$ are matrix of coefficients for $R_{i-m}$ and $E_{i-n}$, respectively. As mentioned in Dufour and Pelletier (2011), we cannot directly work with the representation in (11) because of an identification problem. Consequently, we impose the restrictions on $\Delta_n$ by supposing that the VARMA representation is in diagonal MA form. More specifically, $\Delta_n=\begin{pmatrix}\delta_{11}^{(n)} & 0\\ 0 & \delta_{22}^{(n)}\end{pmatrix}$ where $\delta_{11}^{(n)}$ and $\delta_{22}^{(n)}$ are the coefficients for *n*th-lag error terms of the actual return change and frictionless return change respectively.

Furthermore, we assume the volatility part follows a multivariate GARCH process.

$$\begin{pmatrix}e_i^b\\ e_i^f\end{pmatrix}=H_i^{1/2}z_i$$

where $z_i$ is the bivariate normal distribution that has the following two moments: $E(z_i)=0, Var(z_i)=I_2$. The normality assumption for the error term is supported by backtesting. The other distributions we tried, such as Normal Inverse Gaussian

(NIG), Student-t or Johnson, overestimated the error distribution in our simulation tests.

$H_t$ is the conditional variance matrix for $R_t$ that should be positive definite. To model the dynamic of $H_t$, we use the DCC structure proposed by Engle (2002) in which $H_t$ is decomposed as follows:

$$H_t = D_t C_t D_t$$

*where*:

$$\begin{cases} D_t = diag(\sqrt{h_{11t}}, \sqrt{h_{22t}}), \quad h_{11t} = \sigma_{11t}^2 dur_t^{\gamma_b}, h_{22t} = \sigma_{22t}^2 dur_t^{\gamma_f} \\ C_t = (diagQ_t)^{-1/2} Q_t (diagQ_t)^{-1/2} \end{cases}$$

$$Q_t = (1-\theta_1-\theta_2)Q + \theta_1 \varepsilon_{t-1}\varepsilon_{t-1}^{'} + \theta_2 Q_{t-1}, \quad \varepsilon_{k,t-1} = \frac{e_{k,t-1}}{\sqrt{h_{kk,t-1}}}, \; k = 1,2 \qquad (12)$$

$Q_t$ is the unconditional correlation matrix of $\{\varepsilon_t\}_{t=1}^{N}$, $\sqrt{h_{11t}}$ ($\sqrt{h_{22t}}$) is the conditional variance for actual (frictionless) return change, and $\gamma_b$ ($\gamma_f$) measures the impact of duration on the volatilities of actual and frictionless return changes, respectively.

A useful feature of the DCC model is that it can be estimated by a two-step approach. Engle and Sheppard (2001) show that the likelihood of the DCC model can be written as the sum of two parts: a mean and volatility part, and a correlation part. Even though the estimators from the two-step estimation are not fully efficient, the one iteration of a Newton-Raphson algorithm applied to total likelihood provides asymptotically efficient estimators.

In the DCC framework each series has its own conditional variance. For both actual and frictionless return changes, we adopt a NGARCH($m$,$n$) process as proposed by Engle and Ng (1993) to capture the cluster as well as the asymmetry in volatility. The process can be written as:

$$h_{11,i} = \sigma_{11,i}^2 \cdot dur_i^{\gamma_b}$$
$$\sigma_{11,i}^2 = \omega^b + \sum_{j=1}^{m} \alpha_j^b \sigma_{11,i-m}^2 + \sum_{j=1}^{n} \beta_j^b \sigma_{11,i-j}^2 (\varepsilon_{i-1}^b - \pi^b)^2 \qquad (13)$$

and

$$h_{22,i} = \sigma_{22,i}^2 \cdot dur_i^{\gamma_f}$$
$$\sigma_{22,i}^2 = \omega^f + \sum_{j=1}^{m} \alpha_j^f \sigma_{22,i-m}^2 + \sum_{j=1}^{n} \beta_j^f \sigma_{22,i-j}^2 (\varepsilon_{i-j}^f - \pi^f)^2 \qquad (14)$$

where $\pi^b$ and $\pi^f$ are used to capture the asymmetry in the conditional volatilities. When $\pi^f$ *or* $\pi^b = 0$, the model will become a standard GARCH model, whereas a negative $\pi^f$ *or* $\pi^b$ indicates that a negative shock will cause higher conditional volatilities for the next moment.

Our structure also explicitly introduces the duration dimension in conditional volatilities. The first study of the impact of duration on volatility by Engle (2000) assumes that the impact is linear, that is $h_{11,i} = \sigma_{11,i}^2 \cdot dur_i$. As suggested in Dionne, Duchesne and Pacurar (2009), this modeling for the unit of time might be restrictive for some empirical data for which conditional volatility can depend on duration in a more complicated way. To make the model more general, we follow Dionne, Duchesne and Pacurar (2009) by assuming the exponential form $h_{11,i} = \sigma_{11,i}^2 \cdot dur_i^{\gamma_b}$ and $h_{22,i} = \sigma_{22,i}^2 \cdot dur_i^{\gamma_f}$. When $\gamma_f$ *or* $\gamma_b = 0$, the volatility will become a standard NGARCH process whereas when $\gamma_f$ *or* $\gamma_b = 1$, it transforms to the similar model studied in Engle (2000).

As mentioned above, our model has three uncertainties: variation in duration, price uncertainty and LOB uncertainty. Deriving a closed form of LIVaR$^c$ would be complicated for multi-period forecasting in the presence of three risks, especially for non-regular time duration. Therefore, once the models are estimated, we follow

Christoffersen (2003) and use Monte Carlo simulations to make multi-step forecasting and to test the model's performance.

## 6. Empirical Results

### 6.1 Seasonality adjustment

It is well known that high-frequency data behave very differently from low-frequency data. Table 1 presents the descriptive statistics of raw and deseasonalized duration, frictionless return changes and actual return changes for which various volumes are chosen for the three studied stocks. From Panel A, we can observe that for the entire sample period (July 2010), SAP is the most liquid stock as the average duration is the shortest and the number of observations is the largest. MRK is the least liquid one. Moreover, as the variables are constructed on tick-by-tick frequency, all three stocks have an average of zero and a very small standard deviation for frictionless return changes and actual return changes. All three stocks present high kurtosis due to the fact that most of the observations are concentrated on their average and co-exist with some extreme values. In addition, the raw data are also characterized by extremely high autocorrelation for both first and second moments for all of the variables.

High-frequency data are characterized by seasonality, which should be cleaned out before estimating any model. To do so, several approaches have been proposed in the literature: Andersen and Bollerslev (1997) use Fourier Flexible Functional (FFF) form to take off seasonality, Dufour and Engle (2000) remove seasonality by applying a simple linear regression with a dummy, and Bauwens and Giot (2000) take off seasonality by averaging over a moving window and linear interpolation.

[Insert Table 1 here]

However, as found in Anatolyev and Shakin (2007) and Dionne, Duchesne and Pacurar (2009), the high-frequency data could behave differently throughout the day as well as between different trading days. Therefore, in order to fully account

for the deterministic part exhibited in data, we apply a two-step deseasonalisation procedure, interday and intraday. Besides, it should also be noted that there exists an open auction effect in our continuous trading dataset similar to the one mentioned by Engle and Russell (1998). More precisely, for each trading day, the continuous trading follows the open auction in which specialists set a price in order to maximize the volume. Once the open auction is finished, the transactions are recorded. Consequently, the beginning of continuous trading is contaminated by extremely short durations. In addition, these short durations could produce negative seasonality factors of duration that are based on previous observations and cubic splines. To address this problem, the data for the first half hour are only used to compute the seasonality factor and then discarded.

The interday trend is extracted under a multiplicative form:

$$dur_{i,inter} = \frac{dur_{i,s}}{\overline{dur_s}}, \quad r^f_{i,inter} = \frac{r^f_{i,s}}{\sqrt{\overline{(r^f_s)^2}}}, \quad r^b_{i,inter} = \frac{r^b_{i,s}}{\sqrt{\overline{(r^b_s)^2}}} \tag{15}$$

where $dur_{i,s}$, $r^f_{i,s}$, $r^b_{i,s}$ are the $i$th duration, frictionless and actual return change for day $s$, respectively and $\overline{dur_s}$, $\overline{(r^f_s)^2}$, $\overline{(r^b_s)^2}$ are the daily average for day $s$ for duration, squared frictionless, and actual return changes, respectively.

Based on inter-day deseasonalized data, the intra-day seasonality is removed by following Engle and Russell (1998):

$$dur_{i,intra} = \frac{dur_{i,inter}}{\sqrt{E(dur_{i,inter})|I_{i-1}}}, \quad r^f_{i,intra} = \frac{r^f_{i,inter}}{\sqrt{E((r^f_{i,inter})^2)|I_{i-1}}}, r^b_{i,intra} = \frac{r^b_{i,inter}}{\sqrt{E((r^b_{i,inter})^2)|I_{i-1}}} \tag{16}$$

where $E(dur_{i,inter})$, $E((r^f_{i,inter})^2)$ and $E((r^b_{i,inter})^2)$ are the corresponding deseasonality factors constructed by averaging the variables over 30-minute intervals for each day of the week and then applying cubic splines to smooth these 30-minute averages. The same day of week shares the same intra-deseasonality curve.

However, it takes different deseasonality factors according to the moment of transaction. Figure 1 illustrates the evolution of the seasonality factors of RWE for duration, frictionless, and actual return changes when $v = 4000$.[7] It is not surprising to see that the frictionless and actual return changes have similar dynamics for the reason that the actual return changes contain the frictionless return changes. However, the magnitude is different for frictionless and actual return changes. Panel B of Tables 1.1, 1.2 and 1.3 report descriptive statistics of deseasonalized durations, frictionless and actual return changes. The raw frictionless and actual return changes have been normalized to have the mean equal to zero and standard deviation equal to one. However, other statistics such as skewness, kurtosis and auto-correlation are not affected by this normalization process. The high kurtosis and auto-correlation will be captured by the proposed models.

[Insert Figure 1 here]

6.2 Estimation results

We use the model presented in Section 5 to fit SAP, RWE and MRK deseasonalized data. The data cover the first week of July 2010. The data from the second week are used as out-of-sample data to test the model's performance. As previously mentioned, the estimation is realized jointly for frictionless and actual return changes. The likelihood function is maximized using Matlab v7.6.0 with Optimization toolbox.

Tables 2.1, 2.2 and 2.3 report the estimation results for actual return changes for SAP, RWE and MRK for $v = 4000$, 4000 and 1800 shares, respectively. It should be noted that for each stock, the frictionless return changes and actual return changes are governed by the same duration process, which is assumed to be strictly exogenous. The high clustering phenomenon is indicated in deseasonalized data by the Ljung-Box statistic (see Table 1, Panel B). Furthermore, the clustering in duration is confirmed by the Log-ACD model. To better fit the data, we retain a

[7] Results on other volumes for RWE, SAP and MRK are available upon request.

log-ACD (2,1) specification for SAP durations, a Log-ACD(2,1) model for RWE durations and a log-ACD(1,1) model for MRK durations. The Ljung-Box statistic on standardized residuals of duration provides the evidence that the Log-ACD model is capable of removing the high autocorrelation identified in deseasonalized duration data. The Ljung-Box statistic with 15 lags is dramatically reduced to 25.08 for SAP, to 39.7 for RWE and to 21.79 for MRK.

Frictionless and actual return changes of the three stocks are also characterized by a high autocorrelation in level and volatility. Moreover, the Ljung-Box statistics with 15 lags on deseasonalized return change and its volatility reject the independence at any significance level for the three stocks. Taking the model efficiency and parsimony into consideration, a VARMA(4,2)-MGARCH((1,3),(1,3)))) [8] model is retained for SAP, a VARMA(5,1)-MGARCH((1,3),(1,3)) model for RWE, and the specification of VARMA(2,2)-MGARCH((1,3),(1,3)) for MRK. The model adequacy is assessed based on standardized residuals and squared standardized residuals. Taking MRK as an example, the Ljung-Box statistics for standardized residuals and squared standardized residuals of actual return changes computed with 5, 15, 15, 20 lags, respectively, are not significant at the 5% level. The Ljung-Box statistic with 15 lags has been significantly reduced after modeling to 7.72. Similar results are obtained for stocks RWE and SAP.

Regarding the estimated parameters, the sum of coefficients in each individual GARCH model is close to one, indicating a high persistence in volatility. Furthermore, non-zero of $\pi^f$ and $\pi^b$ in both structures provides the evidence of asymmetry, that is, a negative shock generates a higher conditional volatility for the next moment. It also should be noted that $\gamma_b$ and $\gamma_f$ are both positive for the three stocks. This means that a longer duration will generate a higher volatility for both actual return and frictionless return changes. In addition, due to the fractional

[8] NGARCH (1,3) for actual return changes and NGARCH(1,3) for frictionless return changes.

exponent, the volatility increases with a decreasing speed when duration becomes longer. It should be noted that in our model the volatility is the product of no-duration scaled variance and duration factor, and the $\gamma_b$ of actual return changes are higher than $\gamma_f$ of frictionless return changes for the three stocks. This implies that the duration factor has a larger impact on actual returns than frictionless returns.

The use of dynamic conditional correlation is justified by the fact that $\theta_1$ and $\theta_2$ in equation (12) are both significantly different from zero for the three stocks. As expected, the conditional correlation of actual return and frictionless return changes is time-varying. The sum of two parameters around 0.8 confirms the high persistence of conditional correlation.

[Insert Table 2 here]

6.3 Model performance and backtesting

In this section, we present the simulation procedure and backtesting results on simulated deseasonalized and re-seasonalized frictionless and actual return changes. Once the model is estimated on tick-by-tick frequency, we can test the model performance and compute frictionless IVaR$^c$ and LIVaR$^c$ by Monte Carlo simulation. One of the advantages of our method is that once the model is estimated, we can compute the simulated deseasonalized IVaR$^c$ and LIVaR$^c$ for any horizon without re-estimating the model. In addition, we can compute the simulation-based re-seasonalized IVaR$^c$ and LIVaR$^c$ in traditional calendar time using the available seasonal factors.

We choose different time intervals to test the model performance. The interval lengths are: 40, 50, 60, 80, 100, 120 and 140 units of time for more liquid stocks SAP and RWE and 20, 30, 40, 50, 60, 80 and 100 for less liquid stock MRK. As the model is applied to deseasonalized data, the simulated duration is not in calendar units. However, they are related in a proportional way. It should also be noticed

that, according to the trading intensity, the simulated duration does not correspond to the same calendar time interval. For a more liquid stock, the same simulated interval relates to a shorter calendar time interval. For instance, in the case of MRK, the interval-length 50 re-samples the one-week data for 190 intervals and corresponds to 13.42 minutes and the 100 interval-length relates to 95 intervals and corresponds to 26.84 minutes. However, for a more liquid stock such as SAP, 50 interval-length corresponds to 5.45 minutes and 140-interval-length to 15.27 minutes.

The simulations for frictionless and actual return changes are realized as follows:

1) We generate the duration between two consecutive transactions since we assume that the duration process is strongly exogenous.

2) With the simulated duration and estimated coefficients of the VARMA-NGARCH model, we obtain the corresponding return changes.

3) We repeat steps 1 and 2 for 10,000 paths and re-sample the data at each path according to the predetermined interval.

4) For each interval, we compute the corresponding $IVaR^c$ and $LIVaR^c$ at the desired level of confidence. To conduct the backtesting for each given interval, we also need to construct the return changes for original out-of-sample data.

To validate the model, we conduct the Unconditional Coverage and Independence tests by applying Kupiec test (1995) and Christoffersen test (1998). The Kupiec test checks whether the empirical failure rate is statistically different from the failure rate we are testing, whereas the Christoffersen test aims at evaluating the independence aspect of the violations. More specifically, it rejects VaR models that generate clustered violations by estimating a first order Markov chain model on the sequence. Table 3 reports the p-values for the Kupiec and the Christoffersen tests upon simulated data for confidence levels of 95%, 97.5%, 99% and 99.5%. The time interval varies from 5 minutes to 15 minutes for SAP, from 5 minutes to

16.67 minutes for RWE and from 5 minutes to 27 minutes for MRK. Most of p-values are higher than 5% indicating that, in general, the model captures well the distribution of frictionless and actual return changes.

[Insert Table 3 here]

Since most of trading and risk management decisions are based on the calendar time and raw data, it might be difficult for practitioners to use simulated deseasonalized data to conduct risk management. To this end, we conduct another Monte Carlo simulation that takes into account the time-varying deterministic seasonality factors. The process is similar to that used for simulating deseasonalized data. However, the difference is that we re-introduce the seasonality factors for duration, actual return changes and frictionless return changes. As seasonality factors vary from one day to another, the simulation should take the day of week into account. More precisely, for the first day, simulated durations are converted to a calendar time of that day and then the corresponding timestamp will identify the seasonality factors for actual and frictionless return changes. The simulation process will continue until the corresponding timestamp surpasses the closing time for the underlying day. In the case of a multiple-day simulation, the processes will continue for another day. Based on the simulated re-seasonalized data, we also compute our $IVaR^c$ and $LIVaR^c$ by repeating the same algorithm.

Table 4 presents the backtesting results on the re-seasonalized simulated data. The time interval varies from 5 minutes to 10 minutes for the three stocks and the confidence levels to test are 95%, 97.5%, 99% and 99.5%. Similar to the test results for simulated deseasonalized data, both the Unconditional Coverage test and Independence test suggest that the simulated re-seasonalized data can also provide reliable high-frequency risk measures for all chosen confidence levels over intervals from 5 to 10 minutes.

[Insert Table 4 here]

# 7 Risks for Waiting Cost, Ex-ante Liquidity Risk and Various IVaRs

## 7.1 Risks for waiting cost

As shown in Section 4, the sums of tick-by-tick frictionless return changes and actual return changes over a given interval can be viewed as the waiting costs related to market risk and total risk, which contains market risk and ex-ante liquidity risk. Consequently, the corresponding $IVaR^c$ and $LIVaR^c$ estimate the risk of losses on these waiting costs. Based on simulated re-seasonalized data from previous section that contain the determinist (seasonal factor) and random (error term) elements, we can further investigate the effect of the ex-ante liquidity risk embedded in the open LOB on the total risk. To this end, we define an impact coefficient of ex-ante liquidity risk[9]:

$$\Gamma_{int,v} = \frac{\widehat{LIVaR^c_{int,v}} - \widehat{IVaR^c_{int}}}{\widehat{LIVaR^c_{int,v}}} \qquad (17)$$

As mentioned above, $IVaR^c_{int,t}$ and $LIVaR^c_{int,v,t}$ are the VaRs for frictionless return changes and actual return changes of volume $v$ for the $t$-th interval. As we simulate the data in the tick-by-tick framework, we can compute the $IVaR^c_{int,t}$ and $LIVaR^c_{int,v,t}$ for any desired interval. Accordingly, $\widehat{IVaR^c_{int}}$ and $\widehat{LIVaR^c_{int,v}}$ are the averages of $IVaR^c_{int,t}$ and $LIVaR^c_{int,v,t}$. As a result, $\Gamma_{int,v}$ assesses, on average, the impact of the ex-ante liquidity risk of volume $v$ on total risk for a given interval. Figure 2 shows how the impact coefficients of ex-ante liquidity perform for intervals from 3 minutes to 10 minutes for the three stocks.

[Insert Figure 2 here]

There are two interesting points to stress after observing the plots. First, the curve is globally increasing, that is, in most of the times, the impact coefficient of the ex-ante liquidity increases when interval increases and will finally converge to its long-

[9] The risk measures in equation (17) are computed for confidence level 95%; in practice, other confidence levels can also be used.

run level. It should be noted that the relation of frictionless return changes and actual return changes can be explicitly expressed by $r_i^b = r_i^f + r_i^{LOB}$, where $r_i^{LOB}$ is the volume-dependent LOB return change for $i$-th transaction. Econometrically, as the sum of two ARMA structures is also an ARMA structure (Engel (1984)), the difference of the frictionless return changes and the actual return changes follow implicitly another ARMA structure. Accordingly, the relationship of $IVaR_{int,t}^c$ and $LIVaR_{int,v,t}^c$ can be generally written as:

$$LIVaR_{int,v,t}^c = IVaR_{int,t}^c + LOBIVaR_{int,v,t}^c + Dep_{int,v,t}^{F,LOB} \tag{18}$$

$LOBIVaR_{int,v,t}^c$ measures the risk associated with the open LOB and $Dep_{int,v,t}^{F,LOB}$ presents the dependence between the frictionless return changes and the actual return changes, which can stand for various dependence measures. However, in our specific modeling, $Dep_{int,v,t}^{F,LOB}$ is the covariance between $r_i^f$ and $r_i^{LOB}$. Therefore, the numerator of equation (17) is the sum of $LOBIVaR_{int,v,t}^c$ and $Dep_{int,v,t}^{F,LOB}$. The fact that the curve is globally increasing in time is due to the higher autocorrelation in LOB return changes over time, which are the changes of magnitude in LOB caused by a given ex-ante volume.

The convergence means that, for the long run, the sum of $LOBIVaR_{int,v,t}^c$ and $Dep_{int,v,t}^{F,LOB}$ is proportional to $LIVaR_{int,v,t}^c$. Recall that in a general GARCH framework, the forward multi-step volatility will converge to its unconditional level. In the present study, once the intervals include sufficient ticks for which the volatilities of both LOB return changes and frictionless return changes reach their unconditional levels, the volatilities of the sum of both return changes will increase with the same speed and therefore the impact coefficient of the ex-ante liquidity will converge to its asymptotic level.

Second, it is interesting to observe that the impact coefficients of ex-ante liquidity of RWE stock are negative for a volume of 1000 shares and become positive when

volumes are 2000, 3000 and 4000 shares. A negative impact coefficient of ex-ante liquidity indicates that volatility for the actual return changes is less than that of the frictionless return changes. In other words, the ex-ante liquidity risk embedded in LOB offsets the market risk. This again results from the fact that off-best levels of LOB are more stable than first level. Based on equation (18), when the volume is small, the negative correlation between frictionless return change and LOB return change plays a more important role in determining the sign of the impact coefficients of ex-ante liquidity. However, for a higher ex-ante volume, the risk of LOB return change also increases but faster than its interaction with frictionless return change. Consequently, the impact coefficients of ex-ante liquidity become positive.

### 7.2 High-frequency ex-ante liquidity premium

Based on the tick-by-tick simulation, we can also compute the high-frequency IVaR and LIVaR for frictionless return and actual return. Using IVaR and LIVaR of the same stock, we can further define a relative interval-dependent liquidity premium as follows:

$$\Lambda_{int,v,t} = \frac{LIVaR_{int,v,t} - IVaR_{int,t}}{LIVaR_{int,v,t}} \tag{19}$$

Similar to the liquidity ratio proposed in Giot and Grammig (2006), $IVaR_{int,t}$ and $LIVaR_{int,v,t}$ are the VaR measures for frictionless return and actual return at the end of $t$-th interval and $int$ is the predetermined interval such as 5-min, 10-min, etc. It should be noticed that our defined actual return and frictionless return do not have the time-additivity property but the frictionless return changes and actual return changes do. Even though the VaR based on frictionless return changes and actual return changes can be used directly in practice, in some situations, practitioners might want to predict their potential loss on frictionless return or actual return instead of frictionless and actual return changes for a precise calendar time point.

To illustrate how our model can be used to provide the ex-ante risk measure for frictionless return and actual return, we first compute the return changes for frictionless returns and actual returns, then calculate the instantaneous frictionless return and actual return at the beginning of the given interval using equations (3) and (4). Once we know the frictionless return and actual return at the beginning of a given interval, we can obtain the frictionless return and actual return for the end of the interval. Figure 3 illustrates how the frictionless return and actual return at the end of an interval are computed. Figure 4 then illustrates the evolutions of IVaR and LIVaR associated with a large liquidation volume for SAP, RWE and MRK during one out-of-sample day, July 12th, 2010. For the three stocks, the IVaR and LIVaR both present an inverse U shape during the trading day. However, for the more liquid stock SAP, the IVaR and LIVaR are less volatile than those of the less liquid stocks RWE and MRK. It also seems that the total risk is smaller during the middle of day. Nonetheless, it should be noted that the smaller VaR in absolute terms does not mean we should necessarily trade at that moment. The IVaR and LIVaR only provide the estimates of potential loss for a given probability at a precise point in time.

[Insert Figure 3 here]

[Insert Figure 4 here]

In addition, the difference between curves on each graph, which measures the risk associated with ex-ante liquidity, varies with time. This is due to the fact that LOB interacts with trades and also changes during the trading days. Smaller (bigger) difference indicates a deeper (shallower) LOB. More specifically, for the least liquid stock, MRK, the ex-ante liquidity risk is more pronounced even for a relatively smaller quantity of 1800 shares. Regarding the more liquid stocks such as SAP and RWE, the ex-ante liquidity risk premiums are much smaller even for the relatively larger quantities of 4000 shares. This again suggests that the ex-ante liquidity risk

becomes more severe when the liquidation quantity is large and the stock is less liquid.

### 7.3 Comparison of LIVaR and other intraday VaRs

Our proposed IVaR and LIVaR, which are validated by backtesting, allow us to conduct more analysis on ex-ante liquidity and compare them with other high-frequency risk measures in existing literature.

The standard IVaR proposed by Dionne, Duchesne and Pacurar (2009) is based on a transaction price that is similar to the closing price in daily VaR computation. However, the resulting IVaR serves as a measure of potential loss of 'paper value' for a frozen portfolio and it omits the ex-ante liquidity dimension. To some extent, the IVaR accounts for an ex-post liquidity dimension; more specifically, it measures the liquidity already consumed by the market. However, active traders are more concerned with ex-ante liquidity because it is related to their liquidation value. For any trader, the risk related to liquidity is always present and the omission of this liquidity dimension can cause a serious distortion from the observed transaction price especially when the liquidation volume is large.

Another major difference is that before obtaining LIVaR, we should compute $LIVaR^c$ for actual return changes, which gives the potential loss in terms of waiting costs over a predetermined interval. Accordingly, the resulting LIVaR provides a risk measure for actual return at a given point in time while the standard IVaR is based on tick-by-tick log-returns, which have time-additive property. It thus directly gives a risk measure in terms of price for a given interval.

The high-frequency VaR proposed by Giot and Grammig (2006) is constructed on mid-quote price and ex-ante liquidation price over an interval of 10 or 30 minutes. The mid-quote price is usually considered as the efficient market price in market microstructure theory. However, from a practical perspective, the traders can rarely obtain the mid-quote price during their transactions. Therefore, the use of mid-

quote price will underestimate the risk faced by high-frequency traders. Instead of taking the mid-quote price as the frictionless benchmark, we take a more realistic price, best bid price, as the frictionless price for traders who aim at liquidating their stock. Consequently, for active day-traders, our LIVaR can be considered an upper bond of risk measure that provides the maximum $p$-th quantile in absolute value when liquidating a given volume $v$. Figure 5 illustrates the difference in constructing the frictionless returns and actual returns.

[Insert Figure 5 here]

## 8. Conclusion

In this paper, we introduce the ex-ante liquidity dimension in an intraday VaR measure using tick-by-tick data. In order to take the ex-ante liquidity into account, we first reconstruct the LOB for three blue chip stocks actively traded in Deutsche Börse (SAP, RWE and MRK) and then define the tick-by-tick actual return that is the log ratio of ex-ante liquidation price computed from a predetermined volume over the previous best ask price. Correspondingly, the proposed $IVaR^c$ and $LIVaR^c$ are based on the frictionless return changes and actual return changes and relate to the ex-ante loss in terms of actual return. In other words, both risk measures can be considered as the waiting costs associated with market risk and liquidity risk.

In order to model the dynamic of actual return, we use a logACD-VARMA-MGARCH structure that allows for both the irregularly spaced durations between two consecutive transactions and stylized facts in changes of actual return. In this setup, the time dimension is supposed to be strongly exogenous. Once the model is estimated, Monte Carlo simulations are used to make multiple-steps forecasts. More specifically, the log-ACD process generates first the tick-by-tick duration while the VARMA-MGARCH simulates the corresponding conditional tick-by-tick frictionless return changes and actual return changes. The model performance is assessed by using the tests of Kupiec (1995) and Christoffersen (1998) on both simulated deseasonalized and re-seasonalized data. Both tests indicate that our

model can correctly capture the dynamics of frictionless returns and actual returns over various time intervals for confidence levels of 95%, 97.5%, 99% and 99.5%.

Our LIVaR provides a reliable measure of total risk for short horizons. In addition, the simulated data from our model can be easily converted to data in the calendar time. Practically, the potential users of our measure could be the high-frequency traders that need to specify and update their trading strategies within a trading day or the market regulators who aim to track the evolution of market liquidity, and the brokers and clearing houses that need to update their clients' intraday margins.

Future research can continue in several directions. Our study is focused on a single stock ex-ante liquidity risk. A possible alternative is to investigate how the IVaR and LIVaR evolve in the case of a portfolio. In particular, the no-synchronization of the durations between two consecutive transactions for each stock is a challenge. Another direction is to test the role of ex-ante liquidity in different regimes. Our study focuses on the liquidity risk premium in a relatively stable period. It could also be interesting to investigate how the liquidity risk behaves during a crisis period. This study will require a more complicated econometric model to take into account different regimes.

# References

Anatolyev, S., Shakin, D., 2007. Trade intensity in the Russian stock market: dynamics, distribution and determinants. Applied Financial Economics 17, 87-104

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4, 115-158

Angelidis, T., Benos, A., 2006. Liquidity adjusted value-at-risk based on the components of the bid-ask spread. Applied Financial Economics 16, 835

Bangia, A., Diebold, F.X., Schuermann, T., Stroughair, J.D., 1999. Modeling liquidity risk with implications for traditional market risk measurement and management. The Wharton Financial Institutions Center WP 90-06

Bauwens, L., Giot, P., 2000. The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. Annals of Economics and Statistics / Annales d'Économie et de Statistique, 117-149

Beltran-Lopez, H., Giot, P., Grammig, J., 2009. Commonalities in the order book. Financial Markets and Portfolio Management 23, 209-242

Bilodeau, Y., 2013. XetraParser [computer software]. HEC Montréal

Black, F., 1971. Toward a fully automated stock exchange. Financial Analysts Journal 27, 28-44

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841-862

Christoffersen, P., 2003. Elements of financial risk management. Academic Press, San Diego, 214

Coppejans, M., Domowitz, I., Madhavan, A., 2004. Resiliency in an automated auction. Working Paper ITG Group

Demsetz, H., 1968. The cost of transacting. Quarterly Journal of Economics 82, 33-53

Dionne, G., Duchesne, P., Pacurar, M., 2009. Intraday Value at Risk (IVaR) using tick-by-tick data with application to the Toronto stock exchange. Journal of Empirical Finance 16, 777-792

Domowitz, I., Hansch, O., Wang, X., 2005. Liquidity commonality and return co-movement. Journal of Financial Markets 8, 351-376

Dufour, A., Engle, R.F., 2000. Time and the price impact of a trade. The Journal of Finance 55, 2467-2498

Dufour, J.-M., Pelletier, D., 2011. Practical methods for modelling weak VARMA processes: Identification, estimation and specification with a macroeconomic application. Discussion Paper McGill University (Department of Economics), CIREQ and CIRANO

Engel, E.M.R.A., 1984. A unified approach to the study of sums, products, time-aggregation and other functions of ARMA processes. Journal of Time Series Analysis 5, 159-171

Engle, R.F., 2000. The econometrics of ultra-high-frequency data. Econometrica 68, 1-22

Engle, R.F., 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics 20, 339-350

Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. The Journal of Finance 48, 1749-1778

Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66, 1127-1162

Engle, R.F., Sheppard, K., 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. NBER Working Papers

Ghysels, E., Jasiak, J., 1998. GARCH for irregularly spaced financial data: The ACD-GARCH model. Studies in Nonlinear Dynamics & Econometrics 2, 133

Giot, P., Grammig, J., 2006. How large is liquidity risk in an automated auction market? Empirical Economics 30, 867-887

Glosten, L.R., Harris, L.E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123

Gourieroux, C., Jasiak, J., 2010. Local likelihood density estimation and Value-at-Risk. Journal of Probability and Statistics 2010, 26Pages

Grammig, J., Maurer, K.-O., 2000. Non-monotonic hazard functions and the autoregressive conditional duration model. Econometrics Journal 3, 16-38

Hautsch, N., 2004. Modelling irregularly spaced financial data: theory and practice of dynamic duration models. Springer-Verlag

Irvine, P.J., Benston, G.J., Kandel, E., 2000. Liquidity beyond the inside spread: Measuring and using information in the Limit Order Book. SSRN eLibrary

Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 2, 73-84

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315-1335

Næs, R., Skjeltorp, J.A., 2006. Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market. Journal of Financial Markets 9, 408-432

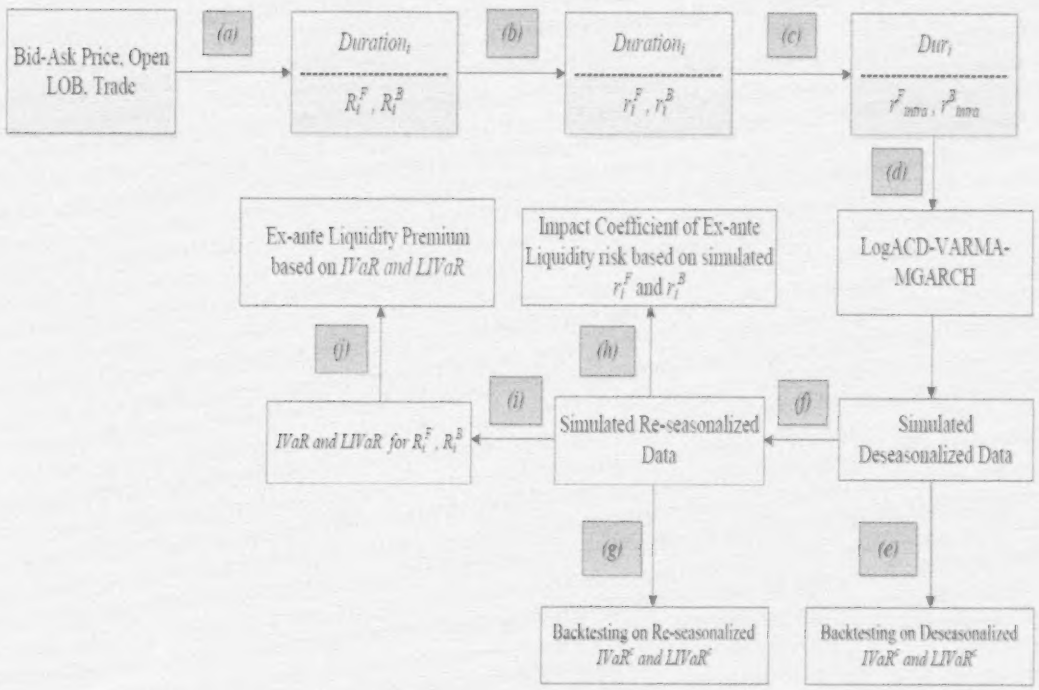
Pacurar, M., 2008. Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. Journal of Economic Surveys 22, 711-751

Pascual, R., Veredas, D., 2010. Does the open limit order book matter in explaining informational volatility? Journal of Financial Econometrics 8, 57-87

Weiß, G.N.F., Supper, H., 2013. Forecasting liquidity-adjusted intraday Value-at-Risk with vine copulas. Journal of Banking & Finance 37, 3334-3350

Zhang, M.Y., Russell, J.R., Tsay, R.S., 2001. A nonlinear autoregressive conditional duration model with applications to financial transaction data. Journal of Econometrics 104, 179-207

## Flowchart for Computing Impact Coefficient of Ex-ante Liquidity Risk and Ex-ante Liquidity Premium




(a): Compute durations, frictionless returns and actual returns based on raw data.

(b): Construct the frictionless return changes and actual return changes.

(c): Remove seasonality from durations, frictionless return changes and actual return changes.

(d): Estimate the model.

(e): Backtest deseasonalized $IVaR^c$ and $LIVaR^c$.

(f): Compute the re-seasonalized data by re-introducing the seasonal factors.

(g): Backtest re-seasonalized $IVaR^c$ and $LIVaR^c$.

(h): Compute the Impact coefficient of ex-ante liquidity risk.

(i): Compute the IVaR and LIVaR for frictionless return and actual return.

(j): Derive the ex-ante liquidity risk premium.

## Table 1.1: Descriptive Statistics for SAP Raw and Deseasonalized Data

The table shows the descriptive statistics for raw durations, actual return changes when Q=2000, 4000, 6000, 8000 and frictionless return changes. The sample period is the first 2 weeks of July 2010 with 44,467 observations.

| Panel A : SAP Raw data | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std.Dev | Skew | Kurt | Min | Max | LB(15) | LB2(15) |
| Duration | 6.43 | 13.21 | 4.58 | 37.28 | 1.00E-03 | 292.75 | 6909.93 | 1426.06 |
| FR Change | 6.39E-09 | 1.93E-04 | 0.10 | 6.01 | -1.98E-03 | 1.47E-03 | 6408.48 | 7082.94 |
| AR(Q=8000) | -7.77E-09 | 1.66E-04 | 0.13 | 8.13 | -1.92E-03 | 2.14E-03 | 5096.10 | 6512.58 |
| AR(Q=6000) | -6.25E-09 | 1.66E-04 | 0.10 | 7.43 | -1.93E-03 | 1.72E-03 | 5252.56 | 6021.21 |
| AR(Q=4000) | -4.73E-09 | 1.68E-04 | 0.09 | 7.79 | -1.96E-03 | 2.11E-03 | 5374.51 | 5185.26 |
| AR(Q=2000) | -2.76E-09 | 1.72E-04 | 0.11 | 7.15 | -1.98E-03 | 1.90E-03 | 5514.05 | 5156.26 |
| Panel B : SAP Deseasonalized Data | | | | | | | | |
| Duration | 1.00 | 1.88 | 4.21 | 37.82 | 0.00 | 52.68 | 4345.54 | 1528.45 |
| FR Change | -1.33E-04 | 1.01E+00 | 0.07 | 5.99 | -10.96 | 7.21 | 6411.40 | 5923.61 |
| AR(Q=8000) | -4.32E-05 | 1.04E+00 | 0.08 | 9.99 | -15.37 | 11.25 | 5043.46 | 6599.80 |
| AR(Q=6000) | -1.05E-05 | 1.04E+00 | 0.08 | 9.45 | -13.07 | 12.60 | 5184.79 | 7189.66 |
| AR(Q=4000) | -6.65E-05 | 1.03E+00 | 0.07 | 8.33 | -12.78 | 10.76 | 5324.84 | 5661.21 |
| AR(Q=2000) | -1.37E-04 | 1.02E+00 | 0.09 | 7.26 | -12.78 | 9.69 | 5501.73 | 4684.47 |

## Table 1.2: Descriptive Statistics for RWE Raw and Deseasonalized Data

The table shows the descriptive statistics for raw durations, actual return changes when Q=1000, 2000, 3000, 4000 and frictionless return changes. The sample period is the first 2 weeks of July 2010 with 37,394 observations.

| Panel A : RWE Raw data | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std.Dev | Skew | Kurt | Min | Max | LB(15) | LB2(15) |
| Duration | 7.64 | 15.88 | 4.75 | 37.53 | 1.00E-03 | 296.58 | 10061.17 | 2479.01 |
| FR Change | -3.10E-08 | 2.51E-04 | 0.12 | 7.51 | -3.07E-03 | 2.49E-03 | 5091.89 | 7971.04 |
| AR(Q=4000) | -4.33E-08 | 2.23E-04 | 0.22 | 10.01 | -2.96E-03 | 2.38E-03 | 3940.66 | 10253.80 |
| AR(Q=3000) | -4.05E-08 | 2.25E-04 | 0.18 | 10.00 | -3.29E-03 | 2.37E-03 | 4022.16 | 10239.86 |
| AR(Q=2000) | -3.70E-08 | 2.27E-04 | 0.18 | 9.19 | -3.22E-03 | 2.33E-03 | 4131.99 | 10306.19 |
| AR(Q=1000) | -3.12E-08 | 2.32E-04 | 0.16 | 8.75 | -3.19E-03 | 2.42E-03 | 4364.74 | 11046.85 |
| Panel B : RWE Deseasonalized Data | | | | | | | | |
| Duration | 0.97 | 1.73 | 3.71 | 24.51 | 2.45E-05 | 28.63 | 3101.10 | 577.62 |
| FR Change | -7.80E-05 | 0.98 | 0.12 | 6.00 | -8.55 | 7.80 | 5123.94 | 5093.30 |
| AR(Q=4000) | -1.23E-04 | 0.99 | 0.20 | 7.32 | -8.61 | 10.11 | 3930.82 | 4781.76 |
| AR(Q=3000) | -1.13E-04 | 0.98 | 0.18 | 7.08 | -9.52 | 9.60 | 4002.22 | 4796.94 |
| AR(Q=2000) | -1.01E-04 | 0.98 | 0.17 | 6.66 | -9.40 | 9.14 | 4101.41 | 4888.15 |
| AR(Q=1000) | -8.61E-05 | 0.98 | 0.17 | 6.38 | -9.27 | 8.52 | 4315.98 | 5519.57 |

# Table 1.3: Descriptive Statistics for MRK Raw and Deseasonalized Data

The table shows the descriptive statistics for raw durations, actual return changes when Q=900, 1800, 2700 and frictionless return changes. The sample period is the first 2 weeks of July 2010 with 17,472 observations.

| Panel A : MRK Raw data | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std.Dev | Skew | Kurt | Min | Max | LB(15) | LB2(15) |
| Duration | 16.31 | 36.40 | 5.16 | 47.06 | 0.001 | 718.35 | 2157.42 | 905.39 |
| FR Change | -3.83E-09 | 3.11E-04 | 0.00 | 15.52 | -5.03E-03 | 4.25E-03 | 2551.10 | 4125.30 |
| AR(Q=2700) | -1.33E-08 | 2.74E-04 | -0.02 | 24.71 | -4.71E-03 | 3.75E-03 | 1672.50 | 2303.16 |
| AR(Q=1800) | -9.58E-09 | 2.79E-04 | 0.17 | 24.70 | -4.77E-03 | 4.74E-03 | 1754.34 | 2308.44 |
| AR(Q=900) | -6.82E-09 | 2.84E-04 | 0.24 | 21.74 | -4.82E-03 | 4.50E-03 | 1890.60 | 3030.64 |
| Panel B : MRK Deseasonalized Data | | | | | | | | |
| Duration | 0.97 | 1.95 | 4.13 | 29.42 | 1.581E-05 | 26.65 | 736.41 | 150.75 |
| FR Change | -2.55E-04 | 1.01 | 0.06 | 13.44 | -13.87 | 11.61 | 2537.61 | 4884.19 |
| AR(Q=2700) | -2.78E-04 | 1.03 | 0.14 | 15.67 | -14.65 | 11.65 | 1670.95 | 2537.68 |
| AR(Q=1800) | -2.81E-04 | 1.03 | 0.20 | 17.60 | -14.15 | 15.63 | 1730.01 | 2201.63 |
| AR(Q=900) | -2.68E-04 | 1.02 | 0.33 | 16.44 | -13.79 | 16.86 | 1862.13 | 2289.72 |

## Table 2.1: Estimation Results SAP ($\nu = 4000$)

The column Statistics reports the Ljung-Box statistic on standardized residuals of duration, actual return changes and squared standardized residuals for different lags. The bold entries are the estimation coefficients that are not significant from zero and the Ljung-Box statistics that reject the no-correlation in the residuals.

| Estimation log-ACD(2,1)-VARMA(4,2)- NGARCH((1,3),(1,3)) (Obs =21089) | | | | | | |
|---|---|---|---|---|---|---|
| | Parameters | Estimation | Std Error | Statistics | | |
| ACD(2,1) | $\alpha_1$ | 0.127 | 0.009 | LB test on Residuals | | |
| | $\alpha_2$ | -0.049 | 0.009 | Lags | Statistic | C_Value |
| | $\beta_1$ | 0.963 | 0.004 | 5 | 5.364 | 11.070 |
| | $\gamma_1$ | 0.812 | 0.022 | 10 | 13.045 | 18.307 |
| | $\gamma_2$ | 0.419 | 0.016 | 15 | **25.077** | 24.996 |
| | $\omega$ | -0.081 | 0.004 | 20 | 30.561 | 31.410 |
| Actual Return Changes VARMA(4,2)-NGARCH((1,3),(1,3)) | $\varphi_{11}^{(1)}$ | 1.043 | 0.088 | LB test on Residuals | | |
| | $\varphi_{12}^{(1)}$ | -0.322 | 0.010 | | | |
| | $\varphi_{11}^{(2)}$ | -0.154 | 0.061 | Lags | Statistic | C_Value |
| | $\varphi_{12}^{(2)}$ | 0.089 | 0.031 | 5 | 5.291 | 11.070 |
| | $\varphi_{11}^{(3)}$ | -0.031 | 0.016 | 10 | 13.424 | 18.307 |
| | $\varphi_{12}^{(3)}$ | 0.047 | 0.014 | 15 | 16.106 | 24.996 |
| | $\varphi_{11}^{(4)}$ | 0.032 | 0.011 | 20 | 18.769 | 31.410 |
| | $\varphi_{12}^{(4)}$ | **0.008** | **0.012** | LB test on Squared Residuals | | |
| | $\delta_{11}^{(1)}$ | -1.338 | 0.087 | | | |
| | $\delta_{11}^{(2)}$ | 0.353 | 0.085 | Lags | Statistic | C_Value |
| | $\omega^b$ | 0.088 | 0.004 | 5 | 3.395 | 11.070 |
| | $\alpha_1^b$ | 0.412 | 0.022 | 10 | 15.405 | 18.307 |
| | $\beta_1^b$ | 0.159 | 0.006 | 15 | 18.925 | 24.996 |
| | $\pi^b$ | 0.404 | 0.022 | 20 | 21.266 | 31.410 |
| | $\alpha_2^b$ | 0.308 | 0.018 | | | |
| | $\gamma^b$ | 0.073 | 0.002 | | | |
| Frictionless Return Changes ARMA(4,2)-GARCH((1,3),(1,3)) | $\varphi_{22}^{(1)}$ | 0.6055 | 0.0693 | LB test on Residuals | | |
| | $\varphi_{21}^{(1)}$ | 0.1754 | 0.0093 | | | |
| | $\varphi_{22}^{(2)}$ | -0.0401 | 0.0175 | Lags | Statistic | C_Value |
| | $\varphi_{21}^{(2)}$ | -0.0593 | 0.0187 | 5 | **15.811** | 11.070 |
| | $\varphi_{22}^{(3)}$ | 0.0488 | 0.0133 | 10 | **21.046** | 18.307 |
| | $\varphi_{21}^{(3)}$ | -0.0199 | 0.0103 | 15 | 23.203 | 24.996 |
| | $\delta_{22}^{(1)}$ | -1.4122 | 0.0690 | 20 | **32.595** | 31.410 |
| | $\delta_{22}^{(2)}$ | 0.4265 | 0.0669 | LB test on Squared Residuals | | |
| | $\omega^f$ | 0.0716 | 0.0037 | | | |
| | $\alpha_1^f$ | 0.4799 | 0.0305 | Lags | Statistic | C_Value |
| | $\beta_1^f$ | 0.1612 | 0.0074 | 5 | 7.794 | 11.070 |
| | $\pi^f$ | 0.3466 | 0.0254 | 10 | 17.916 | 18.307 |
| | $\alpha_2^f$ | 0.2472 | 0.0237 | 15 | 21.902 | 24.996 |
| | $\gamma^f$ | 0.0377 | 0.0025 | 20 | 24.743 | 31.410 |
| DCC parameter | $\theta_1$ | 0.0791 | 0.0032 | | | |
| | $\theta_2$ | 0.7537 | 0.0113 | | | |

# Table 2.2: Estimation Results RWE ($\nu = 4000$)

The column Statistics reports the Ljung-Box statistic on standardized residuals of duration, actual return changes and squared standardized residuals for different lags. The bold entries are the estimation coefficients that are not significant from zero and the Ljung-Box statistics that reject the no-correlation in the residuals.

| Estimation log-ACD(3,1) VARMA(5,1)- NGARCH((1,3),(1,3)) (Obs =15320) | | | | | | |
|---|---|---|---|---|---|---|
| | Parameter | Estimation | StdError | Statistics | | |
| ACD(3,1) | $\alpha_1$ | 0.108 | 0.010 | LB test on Residuals | | |
| | $\alpha_2$ | -0.020 | 0.013 | | | |
| | $\alpha_3$ | -0.024 | 0.009 | Lags | Statistic | C_Value |
| | $\beta_1$ | 0.970 | 0.004 | 5 | 9.209 | 11.070 |
| | $\gamma_1$ | 1.145 | 0.028 | 10 | **20.280** | 18.307 |
| | $\gamma_2$ | 0.280 | 0.010 | 15 | **39.712** | 24.996 |
| | $\omega$ | -0.068 | 0.004 | 20 | **44.304** | 31.410 |
| Actual Return Changes VARMA(5,1)-NGARCH((1,3),(1,3)) | $\varphi_{11}^{(1)}$ | 0.710 | 0.012 | LB test on Residuals | | |
| | $\varphi_{12}^{(1)}$ | -0.333 | 0.011 | | | |
| | $\varphi_{11}^{(2)}$ | 0.047 | 0.014 | | | |
| | $\varphi_{12}^{(2)}$ | **0.005** | **0.013** | Lags | Statistic | C_Value |
| | $\varphi_{11}^{(3)}$ | **0.006** | **0.014** | 5 | 10.161 | 11.070 |
| | $\varphi_{12}^{(3)}$ | 0.024 | 0.013 | 10 | 14.653 | 18.307 |
| | $\varphi_{11}^{(4)}$ | 0.006 | 0.014 | 15 | **27.042** | 24.996 |
| | $\varphi_{12}^{(4)}$ | **0.009** | **0.013** | 20 | 30.095 | 31.410 |
| | $\varphi_{11}^{(5)}$ | 0.041 | 0.012 | LB test on Squared Residuals | | |
| | $\varphi_{12}^{(5)}$ | **0.007** | **0.012** | | | |
| | $\delta_{11}^{(1)}$ | -0.980 | 0.002 | | | |
| | $\omega^b$ | 0.099 | 0.006 | Lags | Statistic | C_Value |
| | $\alpha_1^b$ | 0.437 | 0.030 | 5 | 4.356 | 11.070 |
| | $\beta_1^b$ | 0.156 | 0.008 | 10 | 8.474 | 18.307 |
| | $\pi^b$ | 0.461 | 0.029 | 15 | 11.071 | 24.996 |
| | $\alpha_3^b$ | 0.260 | 0.023 | 20 | 14.102 | 31.410 |
| | $\gamma^b$ | 0.067 | 0.003 | | | |
| Frictionless Return Changes VARMA(5,1)-NGARCH((1,3),(1,3)) | $\varphi_{22}^{(1)}$ | 0.2324 | 0.0123 | | | |
| | $\varphi_{21}^{(1)}$ | 0.1447 | 0.0109 | LB test on Residuals | | |
| | $\varphi_{22}^{(2)}$ | 0.0374 | 0.0132 | | | |
| | $\varphi_{21}^{(2)}$ | **0.0048** | **0.0131** | Lags | Statistic | C_Value |
| | $\varphi_{22}^{(3)}$ | 0.0746 | 0.0135 | 5 | 6.378 | 11.070 |
| | $\varphi_{21}^{(3)}$ | -0.0428 | 0.0133 | 10 | 9.420 | 18.307 |
| | $\varphi_{22}^{(4)}$ | 0.0305 | 0.0138 | 15 | 13.824 | 24.996 |
| | $\varphi_{21}^{(4)}$ | **-0.0168** | **0.0132** | 20 | 22.560 | 31.410 |
| | $\varphi_{22}^{(5)}$ | **0.0205** | **0.0128** | | | |
| | $\varphi_{21}^{(5)}$ | 0.0179 | 0.0114 | | | |
| | $\delta_{22}^{(1)}$ | -0.9805 | 0.0018 | LB test on Squared Residuals | | |
| | $\omega^f$ | 0.0711 | 0.0049 | | | |
| | $\alpha_1^f$ | 0.4586 | 0.0343 | Lags | Statistic | C_Value |
| | $\beta_1^f$ | 0.1436 | 0.0080 | 5 | 7.403 | 11.070 |
| | $\pi^f$ | 0.4531 | 0.0359 | 10 | 16.284 | 18.307 |
| | $\alpha_3^f$ | 0.2731 | 0.0276 | 15 | 19.750 | 24.996 |
| | $\gamma^f$ | 0.0299 | 0.0029 | 20 | 22.811 | 31.410 |
| DCC parameter | $\theta_1$ | 0.0813 | 0.0043 | | | |
| | $\theta_2$ | 0.7791 | 0.0135 | | | |

# Table 2.3: Estimation Results MRK ($\nu = 1800$)

The column Statistics reports the Ljung-Box statistic on standardized residuals of duration, actual return changes and squared standardized residuals for different lags. The bold entries are the estimation coefficients that are not significant from zero and the Ljung-Box statistics that reject the no-correlation in the residuals.

| Estimation log-ACD(1,1)-VARMA(1,1)-NGARCH((1,3),(1,3)) (Obs =7653) | | | | | | |
|---|---|---|---|---|---|---|
| | | | | Statistics | | |
| | Parameter | Estimation | StdError | LB test on Residuals | | |
| **ACD(1,1)** | $\alpha_1$ | 0.048 | 0.005 | Lags | Statistic | C_Value |
| | $\beta_1$ | 0.984 | 0.004 | 5 | 8.240 | 11.070 |
| | $\gamma_1$ | 0.784 | 0.034 | 10 | 17.239 | 18.307 |
| | $\gamma_2$ | 0.382 | 0.025 | 15 | 21.796 | 24.996 |
| | $\omega$ | -0.049 | 0.005 | 20 | 25.006 | 31.410 |
| **Actual Return Changes VARMA(2,2)-NGARCH((1,3),1)** | $\varphi_{11}^{(1)}$ | 0.9401 | 0.0599 | LB test on Residuals | | |
| | $\varphi_{12}^{(1)}$ | -0.3679 | 0.0149 | Lags | Statistic | C_Value |
| | $\varphi_{11}^{(2)}$ | **-0.0316** | **0.0515** | 5 | 11.971 | 11.070 |
| | $\varphi_{12}^{(2)}$ | 0.0945 | 0.0264 | 10 | 25.105 | 18.307 |
| | $\delta_{11}^{(1)}$ | -1.1788 | 0.0584 | 15 | 27.396 | 24.996 |
| | $\delta_{11}^{(2)}$ | 0.2028 | 0.0565 | 20 | 29.201 | 31.410 |
| | $\omega^b$ | 0.1543 | 0.0109 | LB test on Squared Residuals | | |
| | $\alpha_1^b$ | 0.3863 | 0.0341 | Lags | Statistic | C_Value |
| | $\beta_1^b$ | 0.1877 | 0.0115 | 5 | 1.818 | 11.070 |
| | $\pi^b$ | 0.4853 | 0.0328 | 10 | 3.084 | 18.307 |
| | $\alpha_3^b$ | 0.2238 | 0.0289 | 15 | 7.719 | 24.996 |
| | $\gamma^b$ | 0.0861 | 0.0032 | 20 | 10.320 | 31.410 |
| **Frictionless Return Changes ARMA(1,3)-GARCH((1,3),1)** | $\varphi_{22}^{(1)}$ | 0.4737 | 0.0725 | LB test on Residuals | | |
| | $\varphi_{21}^{(1)}$ | 0.1822 | 0.0129 | Lags | Statistic | C_Value |
| | $\varphi_{22}^{(2)}$ | 0.0422 | 0.0228 | 5 | **20.248** | 11.070 |
| | $\varphi_{21}^{(2)}$ | -0.0652 | 0.0179 | 10 | **28.844** | 18.307 |
| | $\delta_{22}^{(1)}$ | -1.2728 | 0.0712 | 15 | **30.197** | 24.996 |
| | $\delta_{22}^{(2)}$ | 0.3020 | 0.0677 | 20 | **34.977** | 31.410 |
| | $\omega^f$ | 0.1551 | 0.0119 | LB test on Squared Residuals | | |
| | $\alpha_1^f$ | 0.4034 | 0.0378 | Lags | Statistic | C_Value |
| | $\beta_1^f$ | 0.2373 | 0.0163 | 5 | 4.348 | 11.070 |
| | $\alpha_3^f$ | 0.1670 | 0.0272 | 10 | 11.419 | 18.307 |
| | $\gamma^f$ | 0.0525 | 0.0039 | 15 | 22.855 | 24.996 |
| **DCC parameter** | $\theta_1$ | 0.1009 | 0.0027 | 20 | **35.233** | 31.410 |
| | $\theta_2$ | 0.6775 | 0.0159 | | | |

## Table 3: Backtesting on Simulated Deseasonalized Data

Panel A: SAP Out-of-sample Backtesting on Deseasonalized Actual Return Change ($t = 4000$)

| Interval (in units) | Kupiec Test | | | | Christoffersen Test | | | | Nb of Intervals | Time interval in minutes |
|---|---|---|---|---|---|---|---|---|---|---|
| | 5% | 2.50% | 1% | 0.50% | 5% | 2.50% | 1% | 0.50% | | |
| 40 | **0.001** | **0.010** | 0.717 | 0.191 | 0.407 | 0.724 | 0.769 | 0.953 | 585 | 4.36 |
| 50 | 0.097 | 0.133 | 0.160 | 0.321 | 0.302 | 0.669 | 0.896 | 0.948 | 468 | 5.45 |
| 60 | 0.907 | 0.090 | 0.286 | 0.971 | 0.174 | 0.747 | 0.919 | 0.919 | 390 | 6.54 |
| 80 | 0.313 | 0.796 | 0.548 | 0.264 | 0.376 | 0.530 | 0.773 | 0.803 | 292 | 8.73 |
| 100 | 0.078 | 0.062 | 0.320 | 0.872 | 0.709 | 0.852 | 0.926 | 0.926 | 234 | 10.90 |
| 120 | 0.342 | 0.679 | 0.971 | 0.980 | 0.503 | 0.723 | 0.838 | 0.919 | 195 | 13.08 |
| 140 | 0.200 | 0.231 | 0.573 | 0.196 | 0.619 | 0.825 | 0.912 | # | 167 | 15.27 |

Panel B: RWE Out-of-sample Backtesting on Deseasonalized Actual Return Change ($t = 4000$)

| Interval (in units) | Kupiec Test | | | | Christoffersen Test | | | | Nb of Intervals | Time interval in minutes |
|---|---|---|---|---|---|---|---|---|---|---|
| | 5% | 2.50% | 1% | 0.50% | 5% | 2.50% | 1% | 0.50% | | |
| 40 | **0.001** | **0.002** | 0.095 | 0.239 | 0.477 | 0.832 | 0.902 | 0.951 | 535 | 4.77 |
| 50 | 0.134 | 0.589 | 0.891 | 0.923 | 0.313 | 0.557 | 0.783 | 0.891 | 428 | 5.96 |
| 60 | 0.217 | 0.292 | 0.362 | 0.874 | 0.341 | 0.679 | 0.881 | 0.881 | 357 | 7.14 |
| 80 | 0.106 | 0.258 | 0.239 | 0.761 | 0.510 | 0.762 | 0.931 | 0.931 | 267 | 9.55 |
| 100 | 0.109 | 0.877 | 0.577 | 0.126 | 0.591 | 0.661 | 0.811 | 0.811 | 214 | 11.92 |
| 120 | 0.145 | 0.460 | 0.522 | 0.909 | 0.590 | 0.748 | 0.915 | 0.915 | 178 | 14.33 |
| 140 | 0.626 | 0.299 | 0.646 | 0.797 | 0.538 | 0.817 | 0.908 | 0.908 | 153 | 16.67 |

Panel C: MRK Out-of-sample Backtesting on Deseasonalized Actual Return Change ($t = 1800$)

| Interval (in units) | Kupiec Test | | | | Christoffersen Test | | | | Nb of Intervals | Time interval in minutes |
|---|---|---|---|---|---|---|---|---|---|---|
| | 5% | 2.50% | 1% | 0.50% | 5% | 2.50% | 1% | 0.50% | | |
| 20 | **0.006** | 0.120 | 0.384 | 0.699 | 0.430 | 0.647 | 0.845 | 0.845 | 476 | 5.36 |
| 30 | **0.026** | 0.119 | 0.152 | 0.617 | 0.547 | 0.782 | 0.936 | 0.936 | 317 | 8.04 |
| 40 | **0.021** | 0.058 | 0.309 | 0.857 | 0.642 | 0.854 | 0.927 | 0.927 | 238 | 10.71 |
| 50 | 0.213 | 0.384 | 0.942 | 0.959 | 0.530 | 0.756 | 0.836 | 0.918 | 190 | 13.42 |
| 60 | 0.470 | 0.613 | 0.747 | 0.820 | 0.490 | 0.732 | 0.820 | 0.910 | 158 | 16.14 |
| 80 | 0.055 | 0.179 | 0.122 | 0.275 | 0.793 | 0.896 | # | # | 119 | 21.43 |
| 100 | 0.378 | 0.800 | 0.959 | 0.506 | 0.656 | 0.768 | 0.883 | 0.883 | 95 | 26.84 |

The table contains the $p$-values for Kupiec and Christoffersen tests for the stocks SAP, RWE and MRK. Interval is the interval length used for computing the LIVaR. Nb of intervals is the number of intervals for out-of sample analysis and Time interval in minutes are the corresponding calendar times. Bold entries indicate the rejections of the model at 95% confidence level. When the numbers of hits are less than two, the $p$-values are denoted by #.

## Table 4: Backtesting on Simulated Re-seasonalized Data

Panel A: SAP Out-of-sample Backtesting on Raw Actual Return Change ($\nu = 4000$)

| Interval (in mins) | Kupiec Test | | | | Christoffersen Test | | | | Nb of Intervals |
|---|---|---|---|---|---|---|---|---|---|
| | 5% | 2.50% | 1% | 0.50% | 5% | 2.50% | 1% | 0.50% | |
| 5 | 0.445 | 0.420 | 0.091 | **0.017** | 0.612 | 0.444 | 0.130 | 0.675 | 485 |
| 6 | 0.132 | 0.483 | 0.256 | 0.986 | 0.334 | 0.570 | 0.888 | 0.888 | 405 |
| 7 | 0.072 | 0.160 | 0.803 | 0.838 | 0.053 | 0.332 | 0.818 | 0.878 | 345 |
| 8 | 0.588 | 0.566 | 1.000 | 0.663 | 0.297 | 0.651 | 0.841 | # | 300 |
| 9 | 0.670 | 0.923 | 0.209 | 0.220 | 0.500 | 0.132 | 0.697 | 0.832 | 270 |
| 10 | 0.066 | 0.726 | 0.362 | 0.520 | 0.081 | 0.552 | 0.752 | 0.898 | 245 |

Panel B: RWE Out-of-sample Backtesting on Raw Actual Return Change ($\nu = 4000$)

| Interval (in mins) | Kupiec Test | | | | Christoffersen Test | | | | Nb of Intervals |
|---|---|---|---|---|---|---|---|---|---|
| | 5% | 2.50% | 1% | 0.50% | 5% | 2.50% | 1% | 0.50% | |
| 5 | 0.173 | 0.341 | 0.364 | 0.778 | 0.266 | 0.604 | 0.897 | 0.927 | 485 |
| 6 | 0.132 | 0.071 | 0.583 | 0.986 | 0.352 | 0.777 | 0.888 | 0.921 | 405 |
| 7 | 0.570 | 0.563 | 0.772 | 0.379 | 0.259 | 0.618 | 0.791 | 0.851 | 345 |
| 8 | 0.070 | 0.156 | 0.537 | 0.697 | 0.651 | 0.776 | 0.908 | 0.908 | 300 |
| 9 | 0.307 | 0.247 | 0.857 | 0.601 | 0.430 | 0.795 | 0.863 | 0.931 | 270 |
| 10 | **0.043** | 0.156 | 0.733 | 0.175 | 0.647 | 0.856 | 0.856 | 0.856 | 245 |

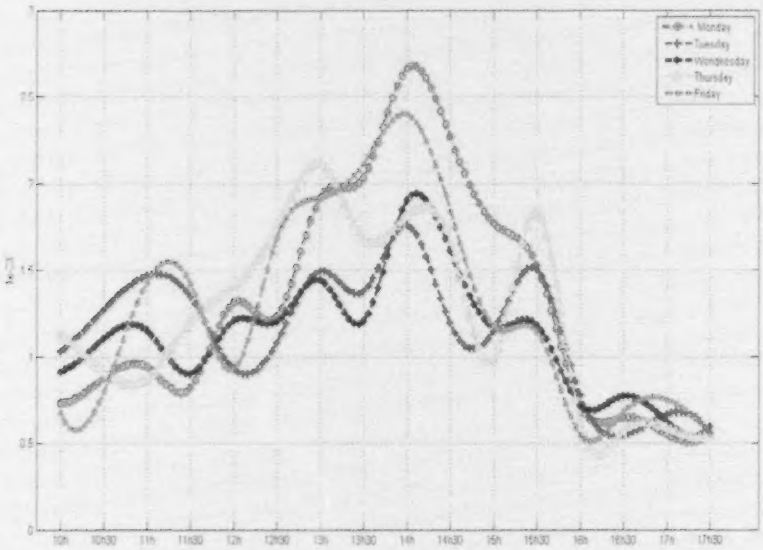
Panel C: MRK Out-of-sample Backtesting on Raw Actual Return ($\nu = 1800$)

| Interval (in mins) | Kupiec Test | | | | Christoffersen Test | | | | Nb of Intervals |
|---|---|---|---|---|---|---|---|---|---|
| | 5% | 2.50% | 1% | 0.50% | 5% | 2.50% | 1% | 0.50% | |
| 5 | 0.187 | 0.201 | 0.140 | 0.298 | 0.344 | 0.628 | 0.897 | 0.949 | 485 |
| 6 | **0.010** | 0.071 | 0.980 | 0.220 | 0.476 | 0.723 | 0.777 | 0.777 | 405 |
| 7 | 0.273 | 0.175 | 0.394 | 0.548 | 0.312 | 0.701 | 0.878 | 0.939 | 345 |
| 8 | 0.267 | 0.566 | 0.537 | 0.083 | 0.359 | 0.620 | 0.870 | # | 300 |
| 9 | 0.098 | 0.247 | 0.654 | 0.752 | 0.513 | 0.728 | 0.863 | 0.931 | 270 |
| 10 | **0.016** | 0.050 | 0.765 | 0.520 | 0.647 | 0.856 | 0.856 | 0.856 | 245 |

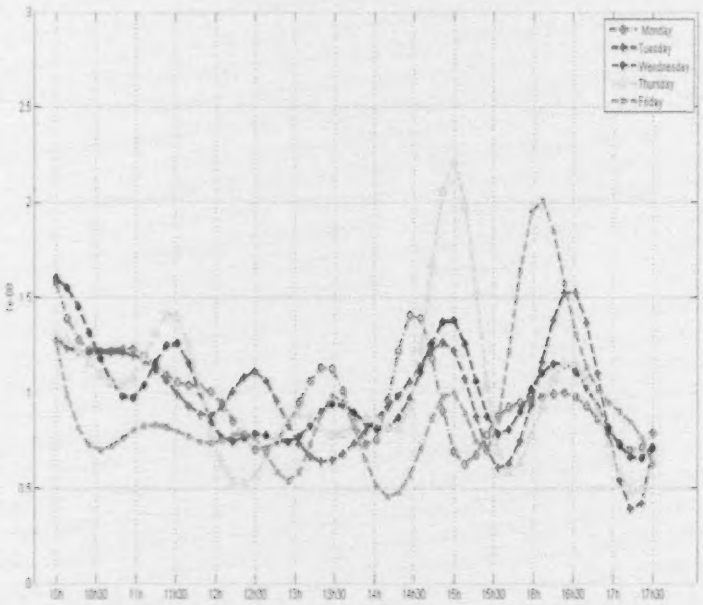
The table contains the $p$-values for Kupiec and Christoffersen tests. Intervals are regularly time-spaced from 5 minutes to 10 minutes. Bold entries indicate the rejections of the model at 95% confidence level. When the numbers of hits are less than two, the $p$-values are denoted by #.

Figure 1: Seasonality Factor For RWE

Panel A: Seasonality Factor for Duration




Panel B: Seasonality Factor for Actual Return Changes ($\nu = 4000$)



Panel C: Seasonality Factor for Frictionless Return Changes

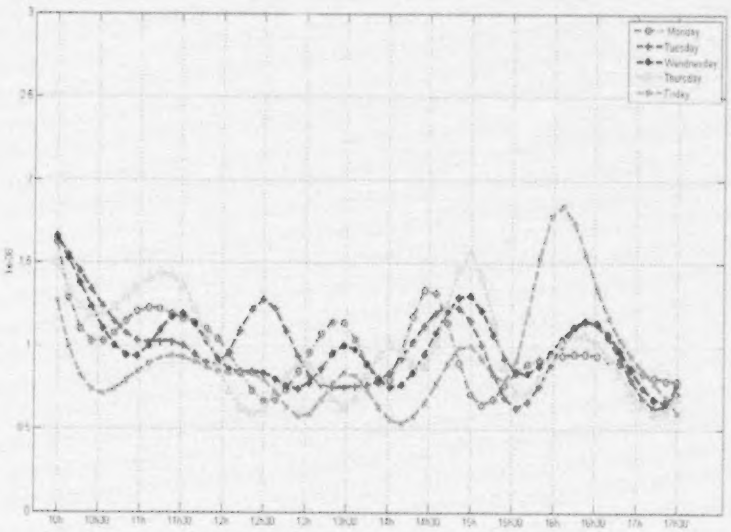
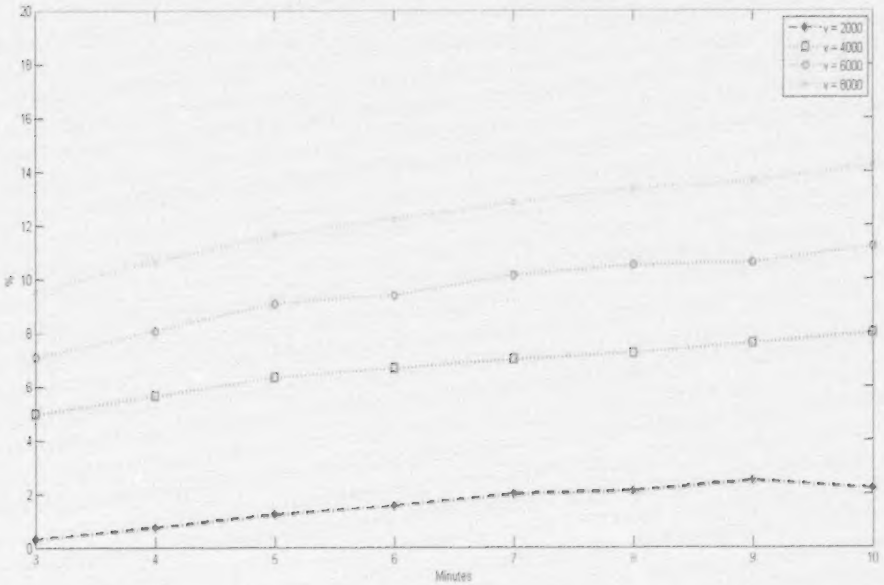
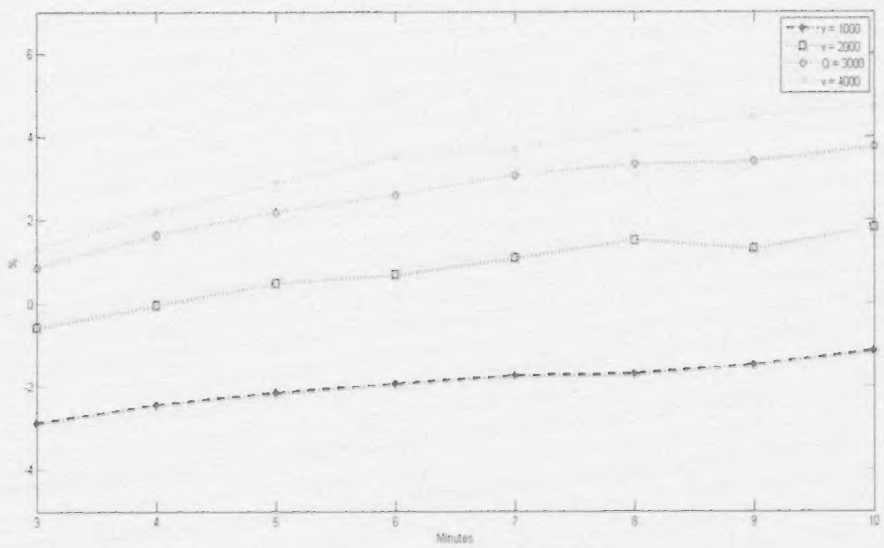


Figure 2: Impact Coefficients of Ex-ante Liquidity for Different Time Intervals

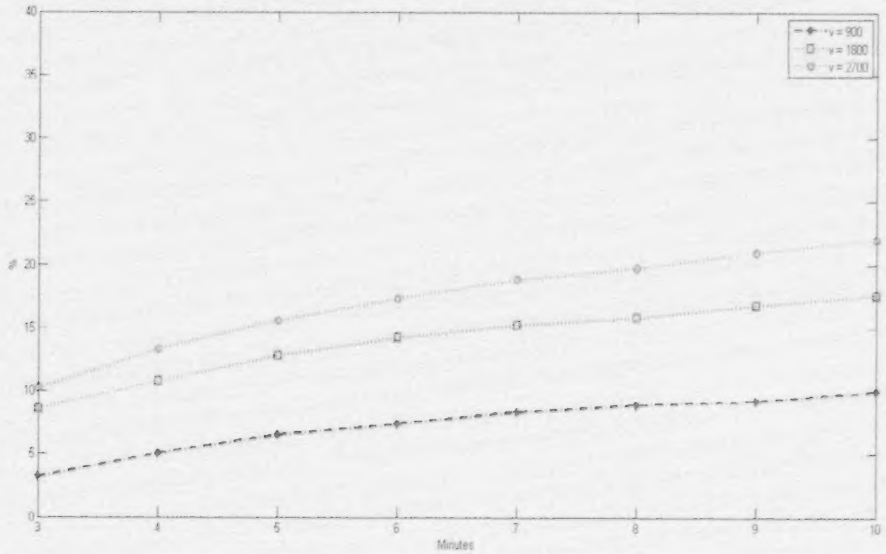
Panel A: SAP




Panel B: RWE



Panel C: MRK



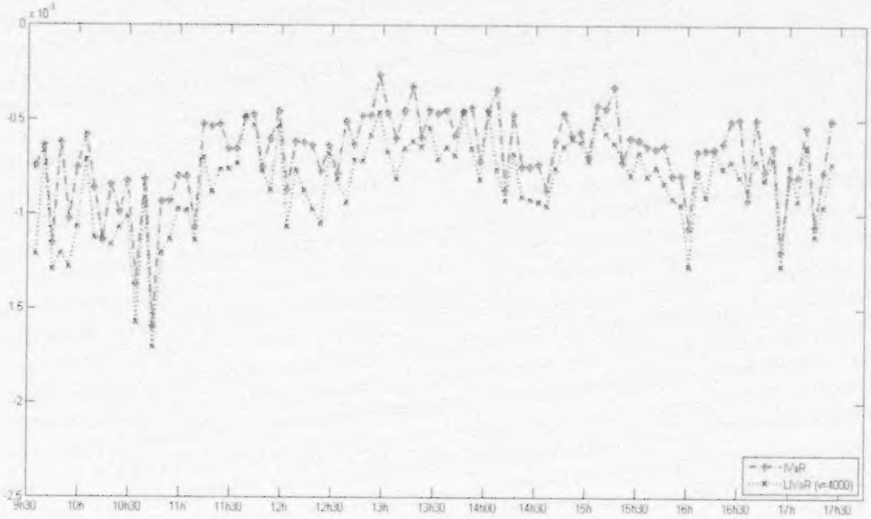

Panels A, B and C illustrate how the impact coefficients of ex-ante liquidity evolve for intervals from 5 minutes to 10 minutes for stocks SAP, RWE and MRK. The selected volumes for the actual return changes are 2000, 4000, 6000, and 8000 shares for SAP, 1000, 2000, 3000, and 4000 shares for RWE, and 900, 1800 and 2700 shares for MRK.

Figure 3: Computation of the Frictionless Return and the Actual Return for the End of an Interval

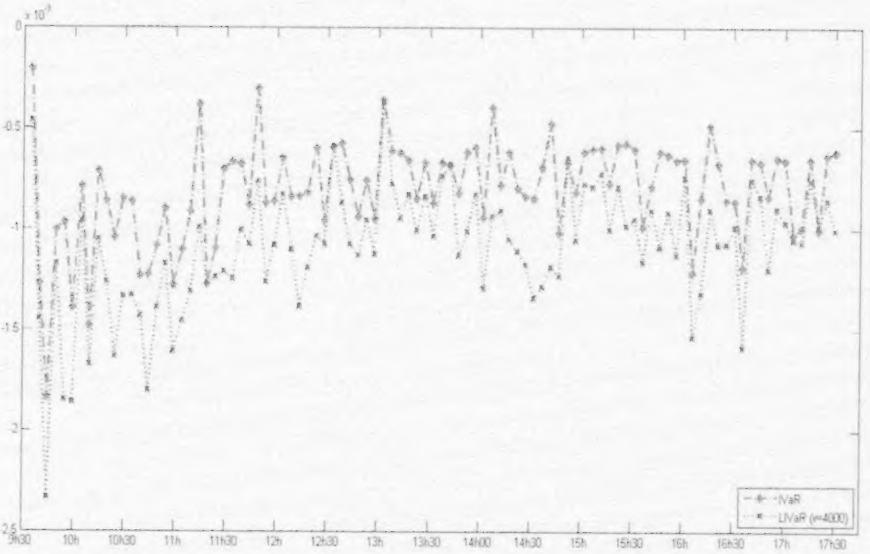
Figure 3 illustrates the computation of the frictionless return and the actual return for the end of an interval. I indicates the transaction. At the beginning of the interval, we compute the frictionless return and actual return using the real data from market. Each I corresponds to a frictionless (actual) change which comes from the simulations. Consequently, the frictionless return (actual return) at the end of the interval is the sum of the initial frictionless return (actual return) and all the corresponding changes in the interval.

Figure 4: IVaR and LIVaR of 5-minute for July 12, 2010

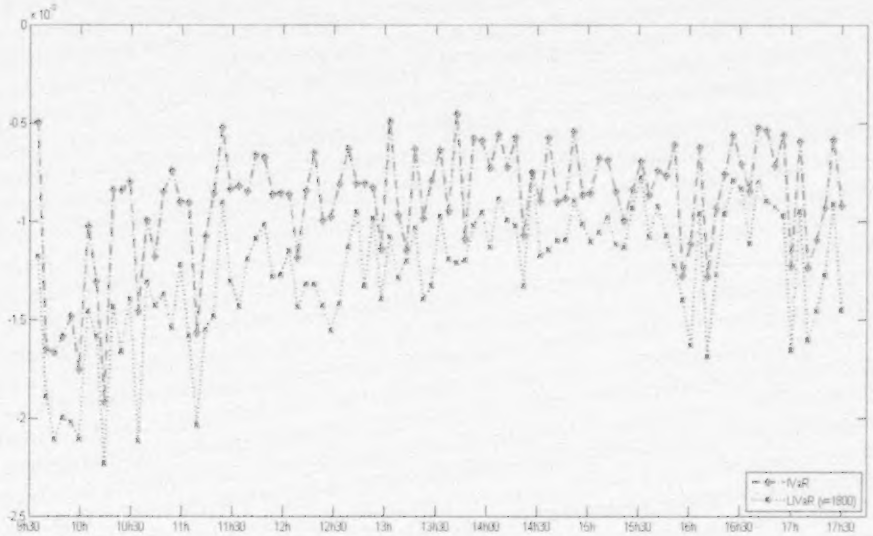
Panel A: SAP with LIVaR ($v$ =4000)




Panel B: RWE with LIVaR ($v$ =4000)

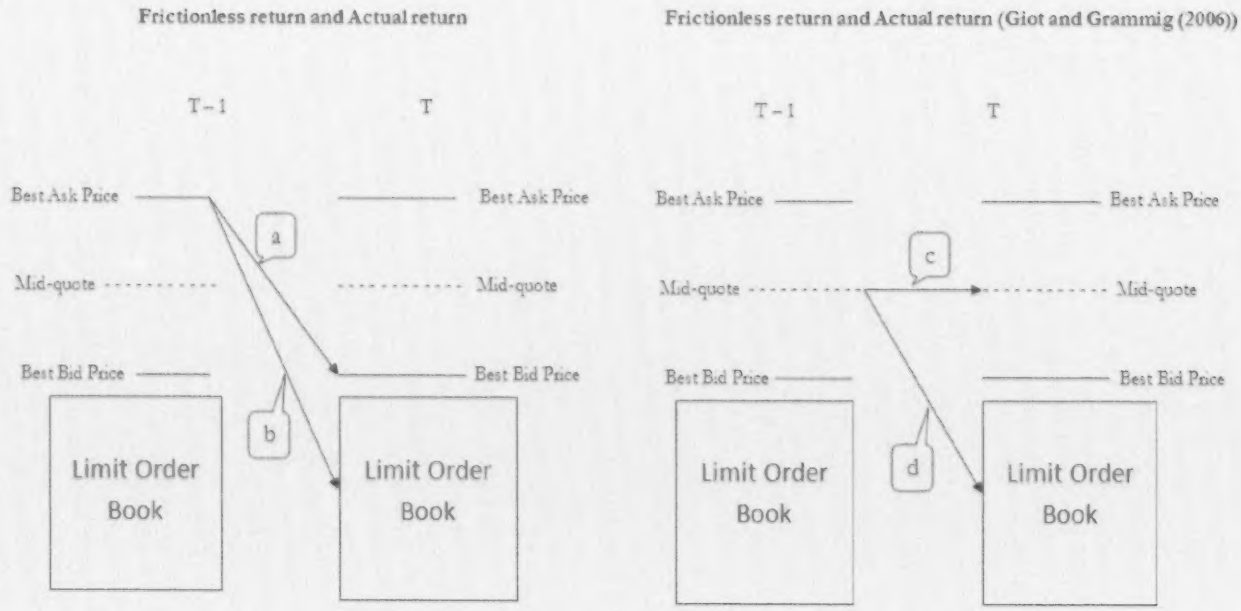


Panel C: MRK with LIVaR ($v$ =1800)




Panels A, B and C present the VaRs for frictionless returns and actual returns at the end of each 5-minute interval on July 12th, 2010, for the three stocks of SAP, RWE and MRK, respectively. The selected volumes for the actual returns are 4000 shares for SAP, 4000 shares for RWE and 1800 shares for MRK.

## Figure 5: Frictionless Returns and Actual Returns




This figure presents the difference between our frictionless (actual) return and the frictionless (actual) return proposed by Giot and Grammig (2006). Arrow *a* presents the starting price and end price in constructing our frictionless return, while arrow *b* shows the starting price and end price for the actual return given a liquidation quantity *v*. Both frictionless returns and actual returns take the previous best ask price as starting price. Arrows *c* and *d* give the starting price and end price for computing the frictionless return and the actual return (for quantity *v*) proposed by Giot and Grammig (2006). Their frictionless return takes the previous mid-quote as the starting price and the following mid-quote as the end price. Their actual return takes the previous mid-quote as the starting price and the actual price as the end price.